Architecture and evolution of the *cis*-regulatory system of the echinoderm *kirrelL* gene

Jian Ming Khor, Jennifer Guerrero-Santoro, and Charles A. Ettensohn*

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA USA

*Corresponding author: Charles A. Ettensohn

Email: ettensohn@cmu.edu

Abstract

The gene regulatory network (GRN) that underlies echinoderm skeletogenesis is a prominent model of GRN architecture and evolution. *KirrelL* is an essential downstream effector gene in this network and encodes an Ig-superfamily protein required for the fusion of skeletogenic cells and the formation of the skeleton. In this study, we dissected the transcriptional control region of the *kirrelL* gene of the purple sea urchin, *Strongylocentrotus purpuratus*. Using plasmid- and BAC-based transgenic reporter assays, we identified key *cis*-regulatory elements (CREs) and transcription factor inputs that regulate *Sp-kirrelL*, including direct, positive inputs from two key transcription factors in the skeletogenic GRN, Alx1 and Ets1. We next identified *kirrelL cis*-regulatory regions from seven other echinoderm species that together represent all classes within the phylum. By introducing these heterologous regulatory regions into developing sea urchin embryos we provide evidence of their remarkable conservation across ~500 million years of evolution. We dissected in detail the *kirrelL* regulatory region of the sea star, *Patiria miniata*, and demonstrated that it also receives direct inputs from Alx1 and Ets1. Our findings identify *kirrelL* as a component of the ancestral echinoderm skeletogenic GRN. They support the view that GRN sub-circuits, including specific transcription factor-CRE interactions, can remain stable over vast periods of evolutionary history. Lastly, our analysis of *kirrelL* establishes direct linkages between a developmental GRN and an effector gene that controls a key morphogenetic cell behavior, cell-cell fusion, providing a paradigm for extending the explanatory power of GRNs.

Introduction

Evolutionary changes in animal form have occurred through modifications to the developmental programs that give rise to anatomy. These developmental programs can be viewed as gene regulatory networks (GRNs, complex, dynamic networks of interacting regulatory (i.e., transcription factor-encoding) genes that determine the transcriptional states of embryonic cells (Peter and Davidson, 2015). Sea urchins and other echinoderms are prominent models for GRN biology for several reasons: (1) there are well-developed tools for dissecting developmental GRNs in these animals, (2) a large number of species that represent a wide range of evolutionary distances are amenable to study, and (3) there is a rich diversity of developmental modes and morphologies within the phylum (Arnone et al., 2016).

All adult echinoderms possess elaborate, calcified endoskeletons. Most species are maximal indirect developers; i.e., they develop via a feeding larva that undergoes metamorphosis to produce the adult. The feeding larvae of echinoids (sea urchins) and ophiuroids (brittle stars) have extensive endoskeletons, holothuroids (sea cucumbers) have rudimentary skeletal elements, and asteroids (sea stars) lack larval skeletal elements entirely. Larval skeletons are thought to be derived within the echinoderms as the feeding larvae of hemichordates (acorn worms), the sister group to echinoderms, and the larvae of crinoids (sea lilies and feather stars), a basal echinoderm clade, lack skeletons. The skeletal cells of larval and adult echinoderms are similar in many respects, supporting the widely accepted view that the larval skeleton arose via co-option of the adult skeletogenic program (Czarkwiani et al., 2013; Gao et al., 2015; Gao and Davidson, 2008; Killian et al., 2010; Mann et al., 2010, 2008; Richardson et al., 1989).

The embryonic skeleton of euechinoid sea urchins, the best studied taxon, is formed by a specialized population of skeletogenic cells known as primary mesenchyme cells (PMCs). These cells are the progeny of the large micromeres (LMs), four cells that arise near the vegetal pole during early cleavage. The GRN that underlies PMC specification is one of the best characterized GRNs in any animal embryo (Oliveri et al., 2008; Shashikant et al., 2018a). This GRN is initially deployed through the activity of a localized maternal protein, Dishevelled, which stabilizes ß-catenin in the LM lineage, leading to the early zygotic expression of a repressor, *pmar1*/*micro1* (Logan et al., 1999; Nishimura et al., 2004; Oliveri et al., 2002; Peng and Wikramanayake, 2013; Weitzel et al., 2004). These molecular events lead to the zygotic expression of several regulatory genes selectively in the LM-PMC lineage. Two of the most important of these regulatory genes are *alx1* (Ettensohn et al., 2003) and *ets1* (Kurokawa et al., 1999), each of which is required for PMC specification and morphogenesis.

After their specification, PMCs undergo a spectacular sequence of morphogenetic behaviors that includes epithelial-mesenchymal transition (EMT), directional cell

migration, cell fusion, and biomineral formation. PMCs undergo EMT at the late blastula stage, ingressing from the vegetal plate into the blastocoel. They migrate along the blastocoel wall and gradually arrange themselves in a ring-like pattern near the equator of the embryo. As they migrate, PMCs extend filopodia that fuse with those of neighboring PMCs, giving rise to a cable-like structure that joins the cells in a single, extensive syncytium. Beginning late in gastrulation and continuing throughout the remainder of embryogenesis, PMCs deposit calcified biomineral within the syncytial filopodial cable.

The complex sequence of PMC morphogenetic behaviors is regulated by hundreds of specialized effector proteins. The spatio-temporal expression patterns of these proteins are controlled by the GRN deployed in the LM-PMC lineage. A major current goal is to identify effector proteins that regulate specific PMC behaviors and elucidate the GRN circuitry that controls these genes (see Ettensohn, 2013; Lyons et al., 2012). Dissection of the *cis*-regulatory elements (CREs) that control essential morphogenetic effector genes, including the identification of specific transcription factor inputs, would directly link them to the relevant circuitry and provide a GRN-level explanation of developmental anatomy. At present, we have only a limited understanding of the *cis*-regulatory control of three PMC effector genes: two genes (*sm30* and *sm50*) that encode secreted proteins occluded in the biomineral (Makabe et al., 1995; Walters et al., 2008) and a third gene (*cyclophilin*/*cyp1*) of unknown function (Amore and Davidson, 2006).

KirrelL is a PMC-specific, Ig domain-containing, transmembrane protein required for cell-cell fusion (Ettensohn and Dey, 2017). In *kirrelL* morphants, PMCs extend filopodia and migrate but filopodial contacts do not result in fusion; this prevents the formation of the PMC syncytium and results in the secretion of small, unconnected biomineralized elements. In all echinoderms that have been examined, the *kirrelL* gene lacks introns, raising the possibility that its origin early in echinoderm evolution was a consequence of retrotransposition, a common gene transfer mechanism that results in intronless genes and one that has played a particularly prominent role in the diversification of Ig-domain-containing proteins (Baertsch et al., 2008; Cordaux and Batzer, 2009; Dermody et al., 2009; Farré et al., 2017). In sea urchins, *kirrelL* is expressed in a temporal and spatial pattern characteristic of many PMC effector genes. The gene is first expressed at the blastula stage (~18 hpf) and peaks in expression early in gastrulation (~30 hpf) (Tu et al., 2014). Expression then declines and is followed by a second peak at ~64 hpf, when *kirrelL* is expressed predominantly at sites of active skeletal rod growth as a consequence of localized, ectoderm-derived cues (Sun and Ettensohn, 2014). *Sp-kirrelL*, like many PMC effector genes, is positively regulated both by Alx1 and Ets1 (Rafiq et al., 2014). Although the gene has only been studied in detail in sea urchins, a recent study found that *kirrelL* is also expressed specifically in the embryonic skeletogenic mesenchyme of a brittle star, *Amphiura filiformis* (Dylus et al., 2018).

In the present study, we used plasmid- and BAC-based transgenic reporter assays to identify key *cis*-regulatory elements (CREs) and transcription factor inputs that regulate *kirrelL* in the sea urchin, *Strongylocentrotus purpuratus*, directly linking this morphogenetic effector gene to the PMC GRN. In addition, we identified *kirrelL cis*-regulatory regions in echinoderm species from all major clades within the phylum and found that these regulatory regions drove PMC-specific expression in developing sea urchin embryos, highlighting their striking conservation across 450-500 million years of evolution. We analyzed in detail the *kirrelL* regulatory region of the sea star, *Patiria miniata*, and found that this gene, like *Sp-kirrelL*, receives direct inputs from Alx1 and Ets1. Our findings identify *kirrelL* as a component of the ancestral echinoderm skeletogenic GRN and strengthen the view that GRN sub-circuits, including specific transcription factor-CRE interactions, can remain stable over very long periods of evolutionary history.

Results

The sea urchin *Sp-kirrelL cis*-regulatory landscape

We identified potential *Sp-kirrelL* CREs based on several criteria. We considered whether candidate regions were (1) hyperaccessible in PMCs relative to other cell types, (2) bound by Alx1, a key transcription factor in the PMC GRN and a positive regulator of *Sp-kirrelL*, (3) associated with active enhnacer RNA (eRNA) expression, and (4) evolutionary conserved. In a previous study, ATAC-seq and DNase-seq were used to identify regions of chromatin that are differentially accessible in PMCs relative to other cell types at the mesenchyme blastula stage (Shashikant et al., 2018b). ChIP-seq was used to indentify binding sites of Sp-Alx1 at the same developmental stage (Khor et al., 2019). Recently, we used Cap Analysis of Gene Expression Sequencing (CAGE-seq) to profile enhancer RNA (eRNA) expression at nine different stages of early sea urchin embryogenesis (Khor et al., 2021). Significantly, our integration of these different genome-wide datasets revealed several putative CREs located near *Sp-kirrelL*, some of which were found to share several signatures **(Figure 1A)**. Developmental CAGE-seq profiles of eRNAs also provided additional information regarding temporal patterns of CRE activity **(Figure 1B)**. To assist in identifying candidate CREs regulating the spatio-temporal expression of *Sp-kirrelL*, we used GenePalette (Smith et al., 2017) to perform phylogenetic footprinting of the *S. purpuratus* and *L. variegatus kirrelL* gene loci. Based on cross-species sequence conservation, cell type-specific DNA accessibility, Sp-Alx1 binding, and eRNA expression, we divided the intergenic sequences flanking *Sp-kirrelL* into 9 putative CREs (labeled elements A-K) **(Figure 1C)**. The elements were between 1.0 to 2.4 kb in size, with an average size of 1.5 kb.

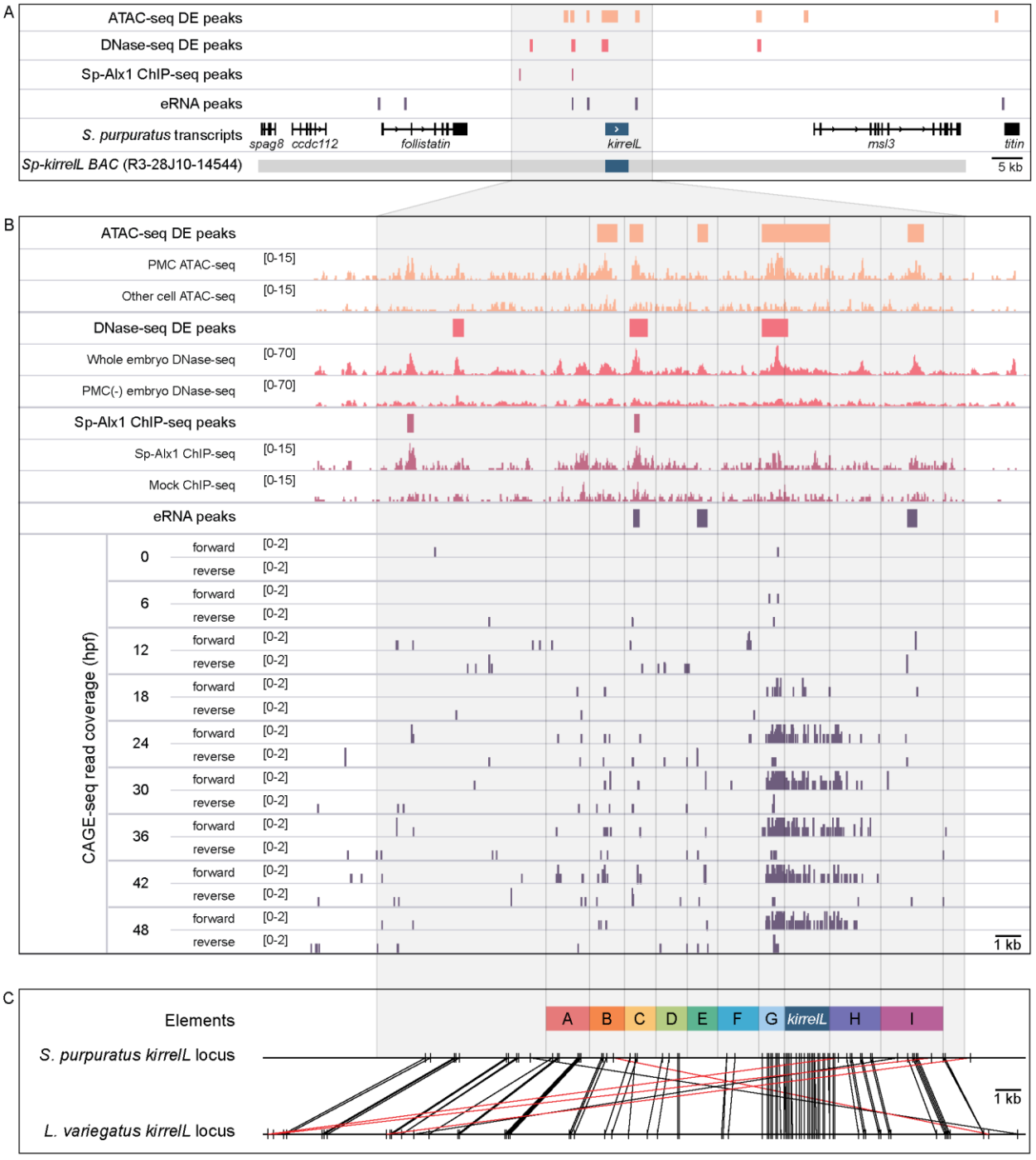


Figure 1: Characterization of the transcriptional regulatory landscape surrounding the *S. purpuratus kirrelL* (*Sp-kirrelL*) locus. (A) Diagram of the *Sp-kirrelL* locus showing neighboring genes, regions of chromatin differentially accessible in primary mesenchyme cells (PMCs) (ATAC-seq DE peaks and DNase-seq DE peaks) (Shashikant et al., 2018), Sp-Alx1 binding sites (Sp-Alx1 ChIP-seq peaks) (Khor et al., 2019), and eRNA peaks (Khor et al., 2021). (B) Signal obtained from each assay in the vicinity of the *Sp-kirrelL* locus. (C) Phylogenetic footprinting of genomic sequences near *S. purpuratus* and *L. variegatus kirrelL* (±10 kb of an exon) using

GenePalette. Black lines indicate identical sequences of 15 bp or longer in the same orientation while red lines indicate identical sequences of 15 bp or longer in the opposite orientation. 9 putative CREs (labeled elements A-I) were identified based on sequence conservation and chromatin signatures.

Characterization of functional *Sp-kirrelL cis*-regulatory elements

To test the transcriptional activity of the candidate *cis*-regulatory elements (**Figure 2A**), we cloned them individually or in combination into the *EpGFPII* reporter plasmid, which contains a weak, basal sea urchin promoter, derived from the *Sp-endo16* gene, upstream of GFP (see Materials and Methods) and injected them into fertilized eggs. We observed that a GFP reporter construct containing upstream elements A to G recapitulated the correct spatial expression pattern of endogenous *Sp-kirrelL* with minimal ectopic expression (**Figure 2B,C and Supplemental Table S1**). Further dissections revealed that a reporter construct containing elements D, E, F, and G also drove strong GFP expression specifically in PMCs while a construct consisting of elements A, B, and C showed weak GFP expression in PMCs. When elements were tested individually, we found that only elements C and G were able to drive GFP expression in sea urchin embryos. Element G, which is directly upstream of the *Sp-kirrelL* translational start site and contains part of the *Sp-kirrelL* 5' untranslated region (UTR), was observed to drive strong GFP expression specifically in the PMCs. Element C was also observed to drive GFP expression specifically in the PMCs, albeit much at lower levels than element G.

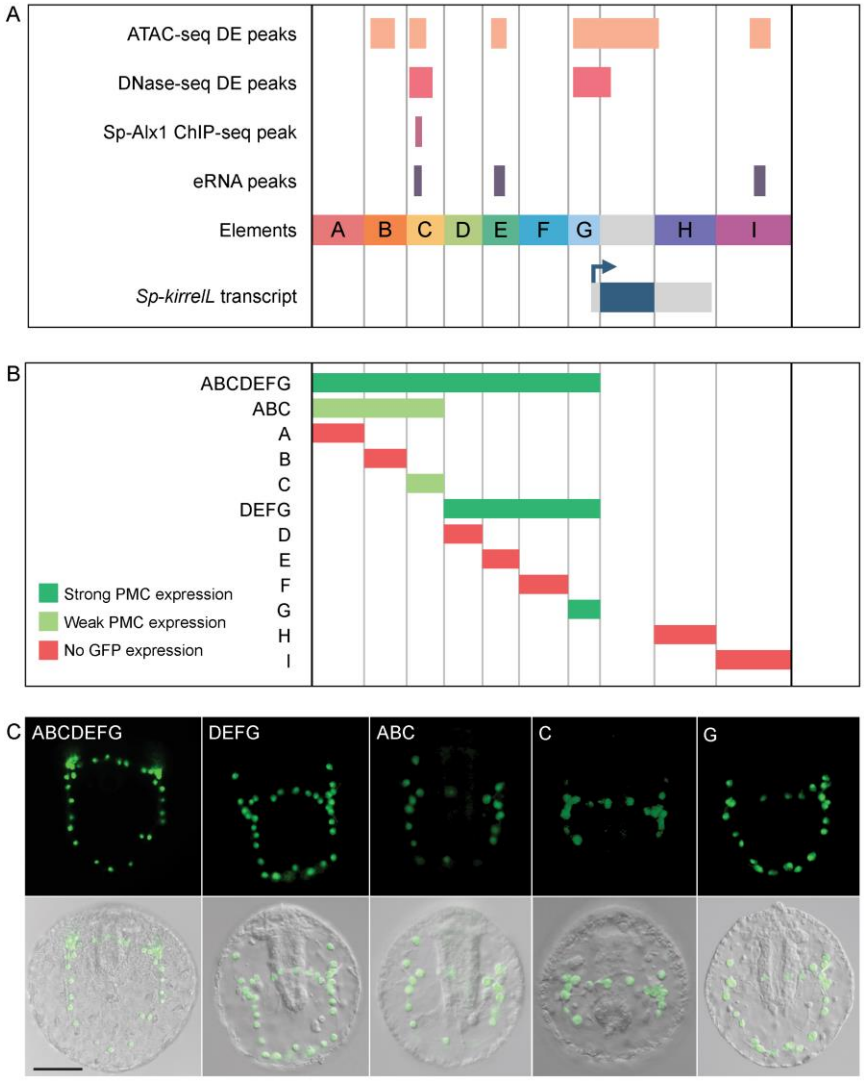


Figure 2: Functional analysis of non-coding genomic sequences flanking *Sp-kirrelL* to identify *cis*-regulatory elements (CREs). (A) 9 putative CREs (labeled elements A-K) were identified based on sequence conservation and previously published datasets (Khor et al., 2021, 2019; Shashikant et al., 2018b). (B) Summary of GFP expression regulated by putative CREs, as assessed by transgenic reporter assays. Strong PMC expression is defined as >15% of total injected embryos expressing GFP, with most exhibiting expression in PMCs only. Weak PMC expression is defined as <15% of total injected embryos expressing GFP, with most exhibiting expression in PMCs only. (C) Spatial expression patterns of GFP reporter constructs containing different *Sp-kirrelL* elements at 48 hours post fertilization (hpf). Top row: GFP fluorescence. Bottom row: GFP fluorescence overlayed onto differential interference contrast (DIC) images. Scale bar: 50 μm.

Identification of direct transcriptional inputs into element C

We next focused on the molecular dissection of element C to identify direct transcriptional inputs into this CRE. Element C is noteworthy as it is differentially accessible in the PMCs based on both ATAC-seq and DNase-seq, bound by Sp-Alx1, and associated with eRNA expression (**Figure 3A**). We first performed a detailed dissection of element C to identify the minimal region that supported strong, PMC-specific GFP expression. We found that a reporter construct containing element C alone showed relatively weak reporter activity, similar to the construct containing elements A, B, and C (**Figure S1A**). In contrast, a larger, overlapping CRE we termed BC.ATAC, which included part, but not all, of element C, exhibited strikingly enhanced GFP expression in PMCs. This difference in activity between element C and BC.ATAC suggested that element C might contain regulatory sites that have greater activity when in close proximity to the promoter.

To explore this further, we generated several reporter constructs consisting of truncated forms of element C, with boundaries defined by peaks from ATAC-seq (C.ATAC), DNase-seq (C.DNase), and Sp-Alx1 ChIP-seq (C.ChIP). The minimal element C region that showed strong, PMC-specific activity was determined to be C.ChIP. Increasing the distance between the C.ChIP element and the promoter (as in the C.DNase construct) significantly reduced enhancer activity. To predict transcription factor inputs within C.ChIP, we scanned the 200 bp C.ChIP sequence using JASPAR (Mathelier et al., 2016), with a focus on transcription factors known to be differentially expressed in the PMCs. This analysis identified several candidate Alx1 and Ets1 binding sites (**Figure 3B and Figure S1B,C**). Consistent with previous RNA-seq analysis which has shown that *Sp-kirrelL* is sensitive to *alx1* and *ets1* knockdowns (Rafiq et al., 2014), our mutational analysis of C.ChIP revealed that mutations of all putative Alx1 and/or Ets1 binding sites completely abolished GFP expression (**Figure 3C and Figure S1D**). In contrast, constructs containing mutations in putative Fox or Meis binding sites exhibited reporter activity similar to that of the parental construct. Mutations of individual Alx1 and Ets1 sites revealed that Alx1 half site 2 and Ets1 site 1 provided key regulatory inputs.

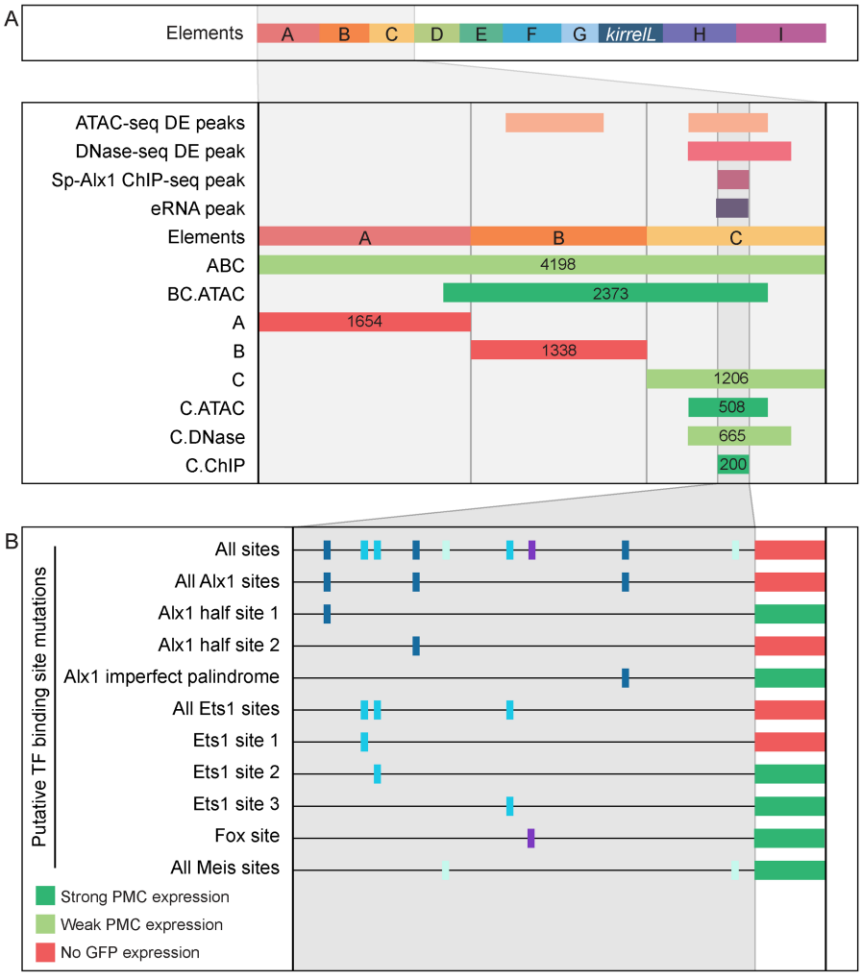


Figure 3: Molecular dissection of element C and the identification of direct transcriptional inputs. (A) Summary of GFP expression regulated by element C truncations using reporter constructs. Serial truncation of element C was performed based on boundaries of peaks defined by chromatin accessibility, Sp-Alx1 binding, and eRNA expression. (B) Summary of GFP expression driven by C.ChIP element mutants. Criteria for strong and weak PMC expression are defined in Figure 2.

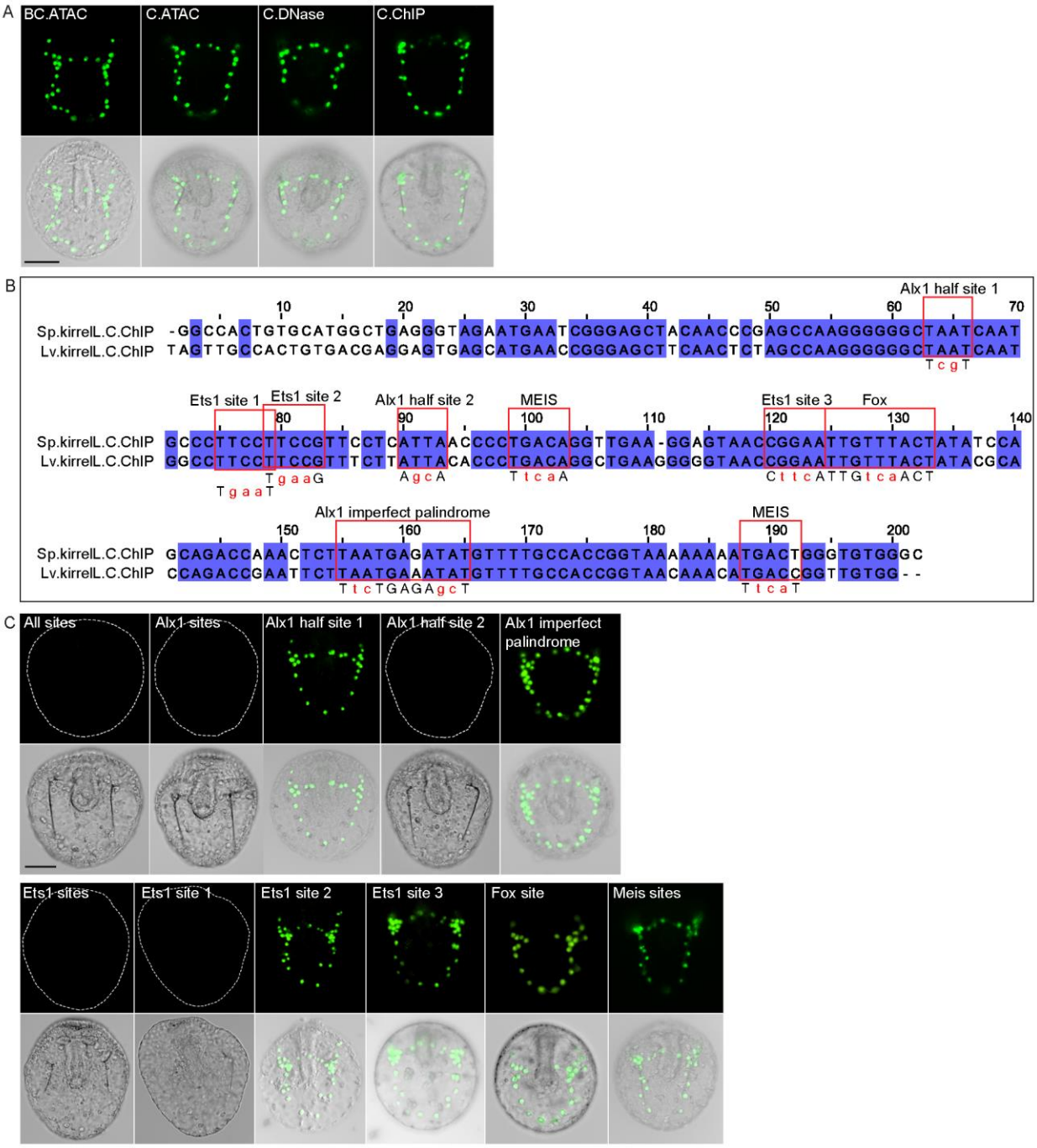


Figure S1: Element C truncation and mutational analysis. (A) Spatial expression patterns of GFP reporter constructs containing different element C truncations at 48 hpf. (B) Cluster Omega alignment of *Sp-kirrelL* and *Lv-kirrelL* C.ChIP sequences. Violet shading indicates conserved sequences. Red boxes highlight putative transcription factor binding sites. (C) Spatial expression patterns of GFP reporter constructs containing different C.ChIP element mutants at 48 hpf. Dotted lines show outline of embryos that did not show GFP expression. Top rows: GFP fluorescence. Bottom rows: GFP fluorescence overlayed onto DIC images. Scale bar: 50 µm.

Analysis of the *Sp-kirrelL* promoter (Element G)

To characterize the native *Sp-kirrelL* promoter, we performed a detailed dissection of element G, which contains sequences directly upstream of the *Sp-kirrelL* translational start site, including the region encoding the *Sp-kirrelL* the 5' UTR (**Figure 4A and Figure S2A**). We found that a 301 bp region surrounding the transcriptional start site, a region we considered to be the *Sp-kirrelL* promoter, was able to drive GFP expression in sea urchin embryos. The majority of expressing embryos, however, exhibited ectopic GFP expression, suggesting that *Sp-kirrelL* core promoter is a strong, ubiquitous promoter. We next performed mutational analysis of the minimal element G fragment that drove strongest PMC-specific GFP expression (G.ATAC). We determined that this CRE receives direct and positive inputs from Alx1 and Ets1, similar to the C.ChIP element (**Figure 4B and Figure S3A,B**). Reporter constructs with mutated Cebpa, Fos::Jun, Fox, Meis, and Tbrain binding sites exhibited similar PMC-specific GFP expression similar to that of the parental construct. We also injected the different *Sp-kirrelL* element G truncations into fertilized *L. variegatus* eggs and observed similar expression patterns, indicating that inputs into element G are conserved in these two sea urchin species (**Figure S2B**).

Our analysis of the native *Sp-kirrelL* promoter prompted us to investigate whether the addition of this region to our *EpGFPII* reporter constructs would allow us to uncover interactions between CREs and the native promoter that would have otherwise been missed (**Figure S4A**). Strikingly, we found that elements B, C, E, F, H, and I were individually able drive strong PMC-specific GFP expression when cloned adjacent to the *Sp-kirrelL* promoter, although these elements had previously exhibited minimal activity in the context of the *Sp-endo16* promoter alone (**Figure 4C and Figure S4B;** compared to **Figure 2**). We also observed that the presence of the native *Sp-kirrelL* promoter mitigated the need for the C.ChIP element within element C to be adjacent to the promoter for strong PMC-specific GFP expression. We confirmed that enhancer activity was dependent on the sequence of the *Sp-kirrelL* promoter, as GFP expression was abolished in a construct where the sequence was shuffled (**Figure S4C,D**). Taken together, these findings showed that several CREs are capable of interacting specifically with the native *Sp-kirrelL* promoter and this interaction can bypass spacing hurdles that are evident when the *Sp-endo16* promoter alone is present.

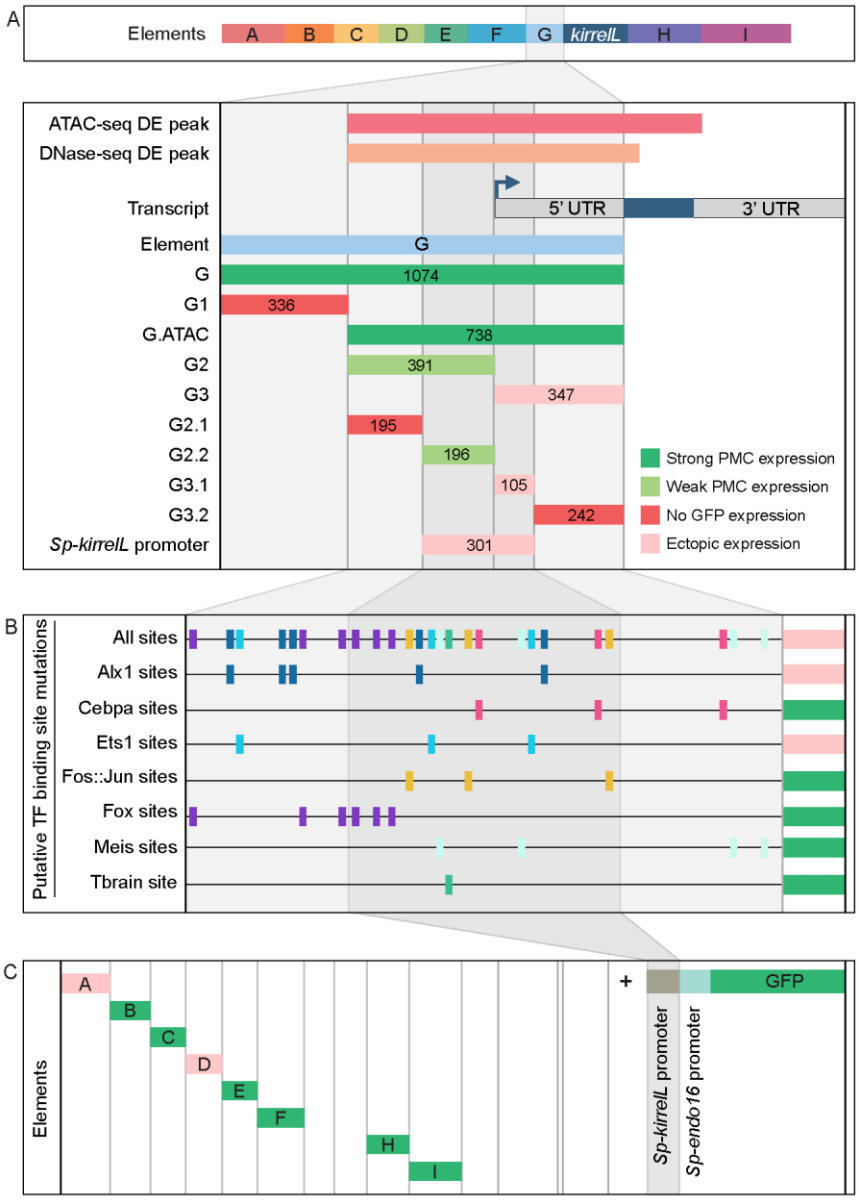


Figure 4: Molecular dissection and mutation of element G. (A) Summary of GFP expression regulated by element G truncations using *EpGFPII* reporter constructs. Serial truncation of element G was performed based on boundaries defined by chromatin accessibility and the *kirrelL* 5' UTR. Criteria for strong and weak PMC expression are defined in Figure 2. Ectopic expression is defined as >10% of total injected embryos expressing GFP, with a majority exhibiting expression in cells other than PMCs. (B) Summary of GFP expression driven by G.ATAC element mutants using *EpGFPII* reporter constructs. (C) Analysis of element enhancer activity in modified *EpGFPII* reporter constructs containing the endogenous *Sp-kirrelL* promoter elements.

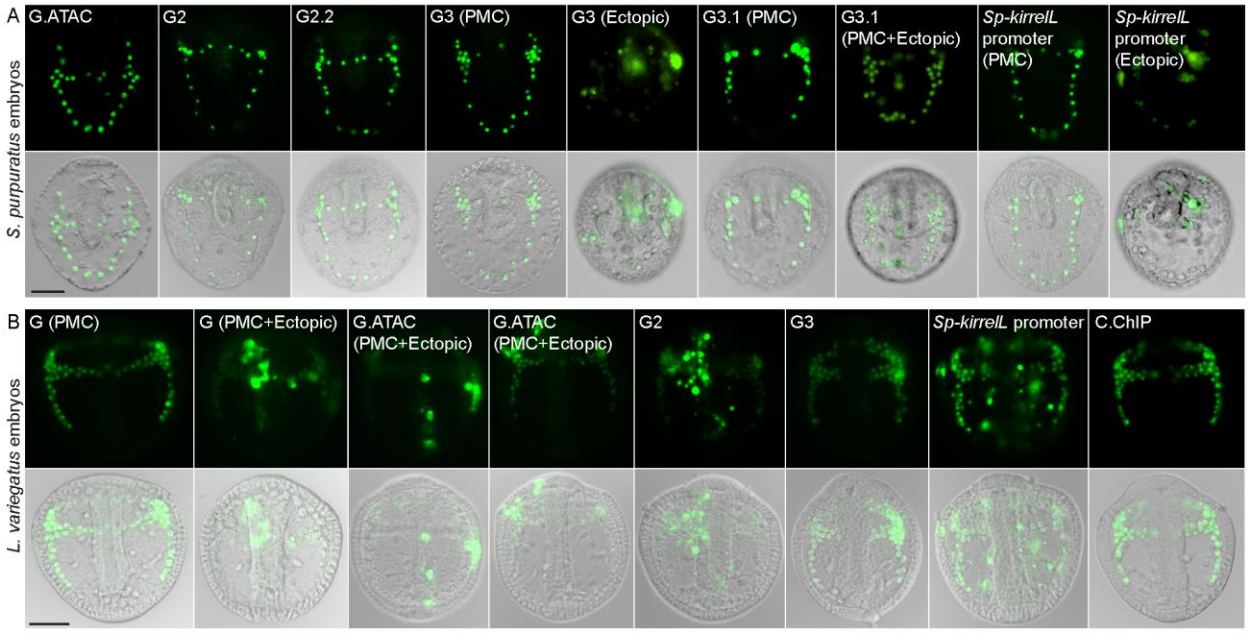


Figure S2: Element G truncation and mutational analysis. (A) Spatial expression patterns of GFP reporter constructs containing element G truncations in *S. purpuratus* embryos at 48 hpf. (B) Spatial expression patterns of GFP reporter constructs containing element G truncations in *L. variegatus* embryos at 28 hpf. Representative images of PMC-specific and ectopic GFP expression are shown for some constructs. Top rows: GFP fluorescence. Bottom rows: GFP fluorescence overlayed onto DIC images. Scale bar: 50 μm.

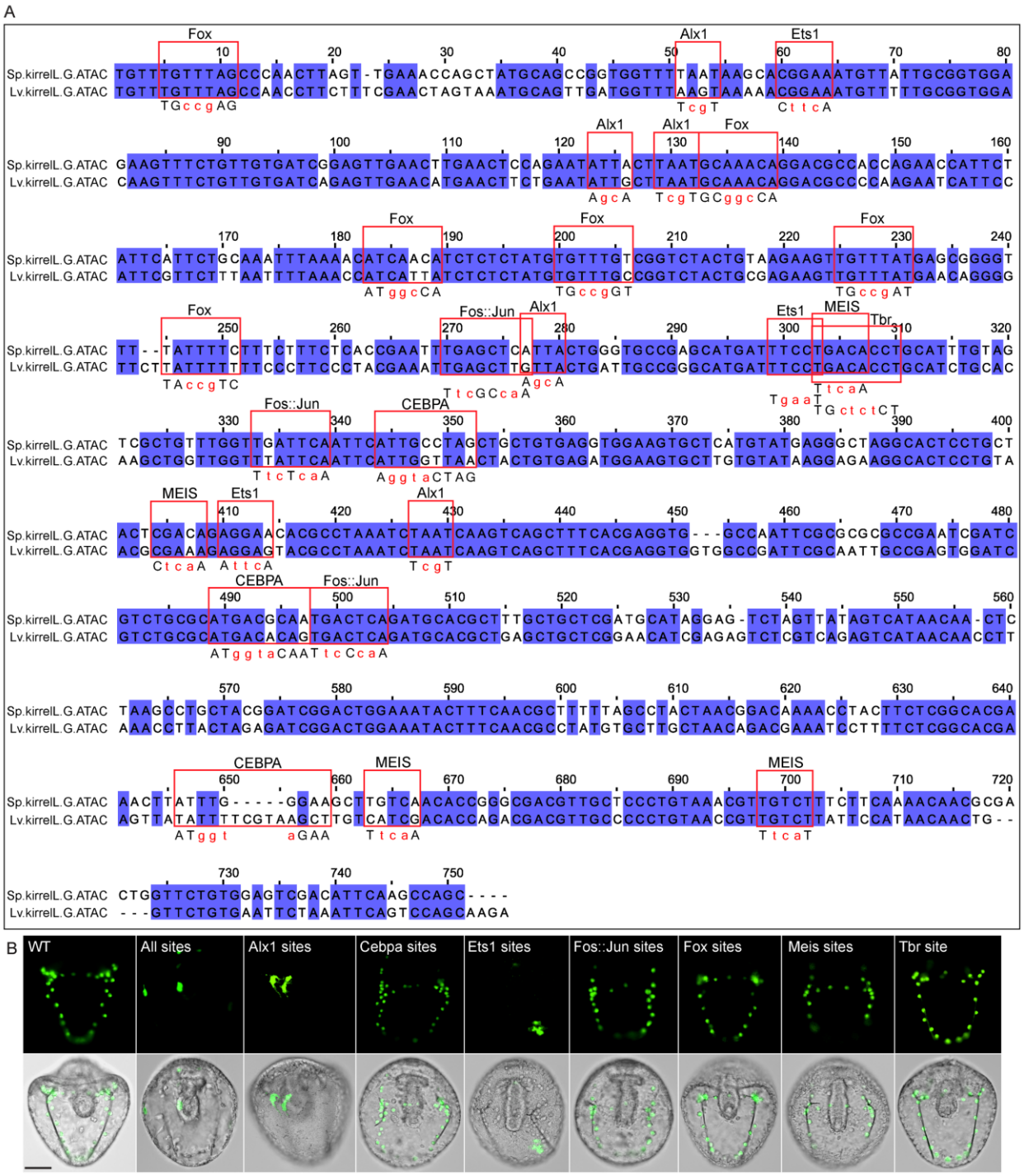
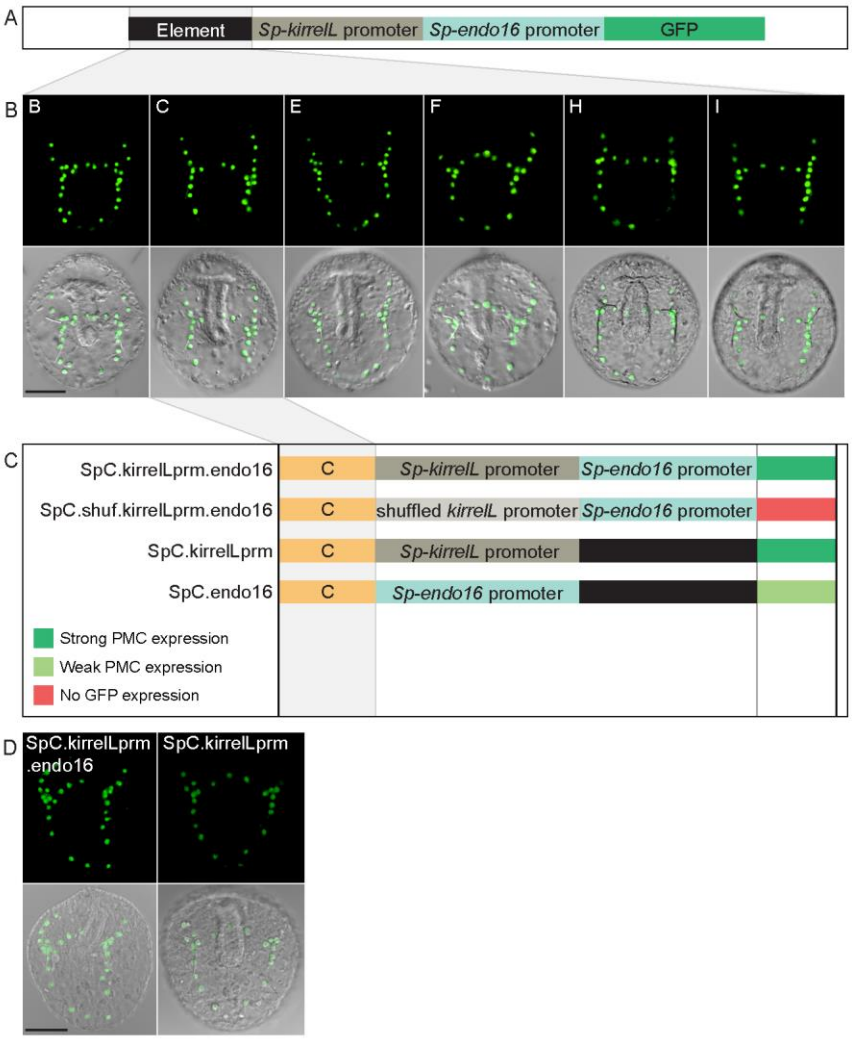


Figure S3: Mutational analysis of G.ATAC element. (A) Cluster Omega alignment of *Sp-kirrelL* and *Lv-kirrelL* G.ATAC sequences. Violet shading indicates conserved sequences. Red boxes highlight putative transcription factor binding sites. (B) Spatial expression pattern of GFP reporter constructs containing different G.ATAC element mutants. Top row: GFP fluorescence. Bottom row: GFP fluorescence overlayed onto DIC images. Scale bar: 50 μm.



299

300 **Figure S4**: Interactions between *Sp-kirrelL* CREs and the endogenous *Sp-kirrelL* promoter. (A)
301 Diagram showing the backbone of a modified, *EpGFPII* reporter construct containing the
302 endogenous *Sp-kirrelL* promoter upstream of the *Sp-endo16* promoter. (B) Spatial expression
303 patterns of the modified GFP reporter constructs containing the various *Sp-kirrelL* elements
304 indicated (elements B, C, E, F, H, and I). (C) Summary of GFP expression patterns of control
305 constructs to determine whether element activity is dependent on the presence of the *Sp-kirrelL*
306 promoter. Criteria for strong and weak PMC expression are defined in Figure 2. (D) Spatial
307 expression patterns of GFP reporter constructs containing element C and the *Sp-kirrelL* promoter,
308 with or without the *Sp-endo16* promoter. Top rows: GFP fluorescence. Bottom rows: GFP
309 fluorescence overlayed onto DIC images. Scale bar: 50 μm.

Relative contributions of individual CREs in the context of the entire *Sp-kirrelL* regulatory apparatus

Our analysis identified multiple CREs in the vicinity of the *Sp-kirrelL* locus that were capable of driving PMC-specific reporter expression when cloned into plasmids that contained the endogenous *Sp-kirrelL* promoter. To explore further the relative contributions of these various elements to *Sp-kirrelL* expression in vivo, we examined their function in the context of the complete transcriptional control system of the gene. For these studies, we utilized a 130 kb bacterial artificial chromosome (BAC) that contained the single exon *Sp-kirrelL* gene, flanked by 65 kb of sequences in each direction. We used recombination-mediated genetic engineering (recombineering) to replace the single *Sp-kirrelL* exon seamlessly with either GFP or mCherry coding sequence (**Figure 5A**). We found that *Sp.kirrelL*.GFP.BAC faithfully recapitulated the expression of endogenous *Sp-kirrelL* in the PMCs at 48 hpf with minimal ectopic expression (**Figure S5**). We next generated deletion mutants based on results from our plasmid GFP reporter assays to quantitatively assess the contributions of elements A to G to *Sp-kirrelL* transcriptional regulation. We found that deletion of elements A to G (ΔCRE.GFP.BAC) completely abolished GFP expression. We also observed that retaining the minimal endogenous *Sp-kirrelL* promoter (ΔCRE.kirrelLprm.GFP.BAC) did not rescue GFP expression, demonstrating that elements A to G are necessary for *Sp-kirrelL* expression in the context of the *Sp.kirrelL*.GFP.BAC consistent with our previous, plasmid-based analysis.

To directly compare the spatial expression patterns of deletion mutants with that of the parental BAC, we generated BAC mutants containing deletion of individual elements and co-injected them into fertilized eggs with a parental mCherry BAC. We found that a BAC containing deletion of the element G (ΔG.GFP.BAC, which included a deletion of the *Sp-kirrelL* promoter) abolished GFP expression at 48 hpf (**Figure 5B and Figure S5**). By contrast, deletion of all of element G except for the promoter region (ΔG.kirrelLpromoter.GFP.BAC), resulted in a GFP spatial expression pattern similar to that of the parental mCherry. These findings confirmed the importance of the *Sp-kirrelL* promoter in supporting PMC-specific expression of the gene and showed that this region is essential even when all distal CREs are present. BACs containing individual deletions of other elements all remained active at 48 hpf and supported PMC-specific reporter expression, pointing to considerable redundancy in the contribution of each element to *Sp-kirrelL* expression.

To examine the relative contribution of distal CREs more rigorously, we measured levels of reporter transcripts using a NanoString nCounter. For each mutant BAC, we co-injected embryos with mCherry tagged, parental BAC and the GFP-tagged, mutant BAC and quantified the expression level of each reporter gene at four time points (20, 30, 50, and 65 hpf) (**Figure 5C and Supplemental Table S2**). We found that deletion of element

C resulted in approximately a 50% reduction in expression compared to WT BAC. As we observed previously, GFP expression was completely abolished when element G was deleted (ΔG.GFP.BAC) and this effect was diminished when the *Sp-kirrelL* promoter was retained (ΔG.kirrelLprm.GFP.BAC). Quantitative analysis revealed, however, that retention of the *Sp-kirrelL* promoter alone resulted in only a partial rescue of expression, with overall levels reduced substaintially compared to the wild-type BAC. We also observed that deletion of element H resulted in decreased expression levels. Taken together, our qualitative and quantitative analyses show that at early stages of embryo development, *Sp-kirrelL* expression is controlled by multiple CREs, notably the C, G, and H modules, acting in concert with the Sp-kirrelL promoter.

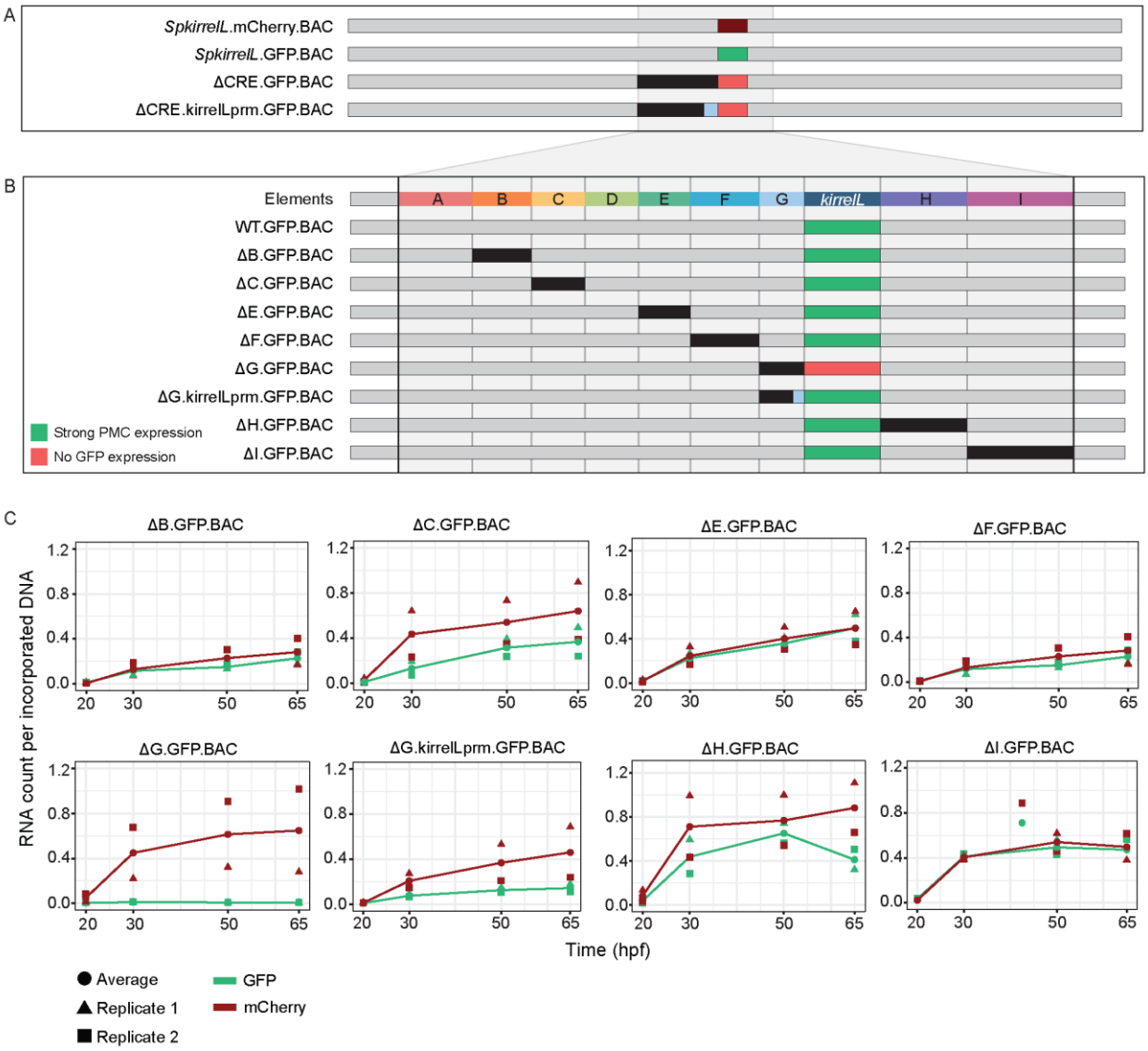


Figure 5: *Sp-kirrelL cis*-regulatory analysis using BACs. (A) BAC deletions show that elements A-G are necessary for GFP expression, regardless of the presence of the endogenous *Sp-kirrelL* core promoter elements. (B) Summary of GFP expression patterns of individual *Sp-kirrelL* elements using GFP BAC deletions. Criteria for strong PMC expression are defined in Figure 2. (C) Quantitative NanoString analysis of reporter expression in embryos co-injected with parental mCherry and mutant GFP BACs. Embryos were collected at 20, 30, 50, and 65 hpf. The average expression profile for each pair of BAC injection was calculated from NanoString counts of two biological replicates (see Materials and Methods).

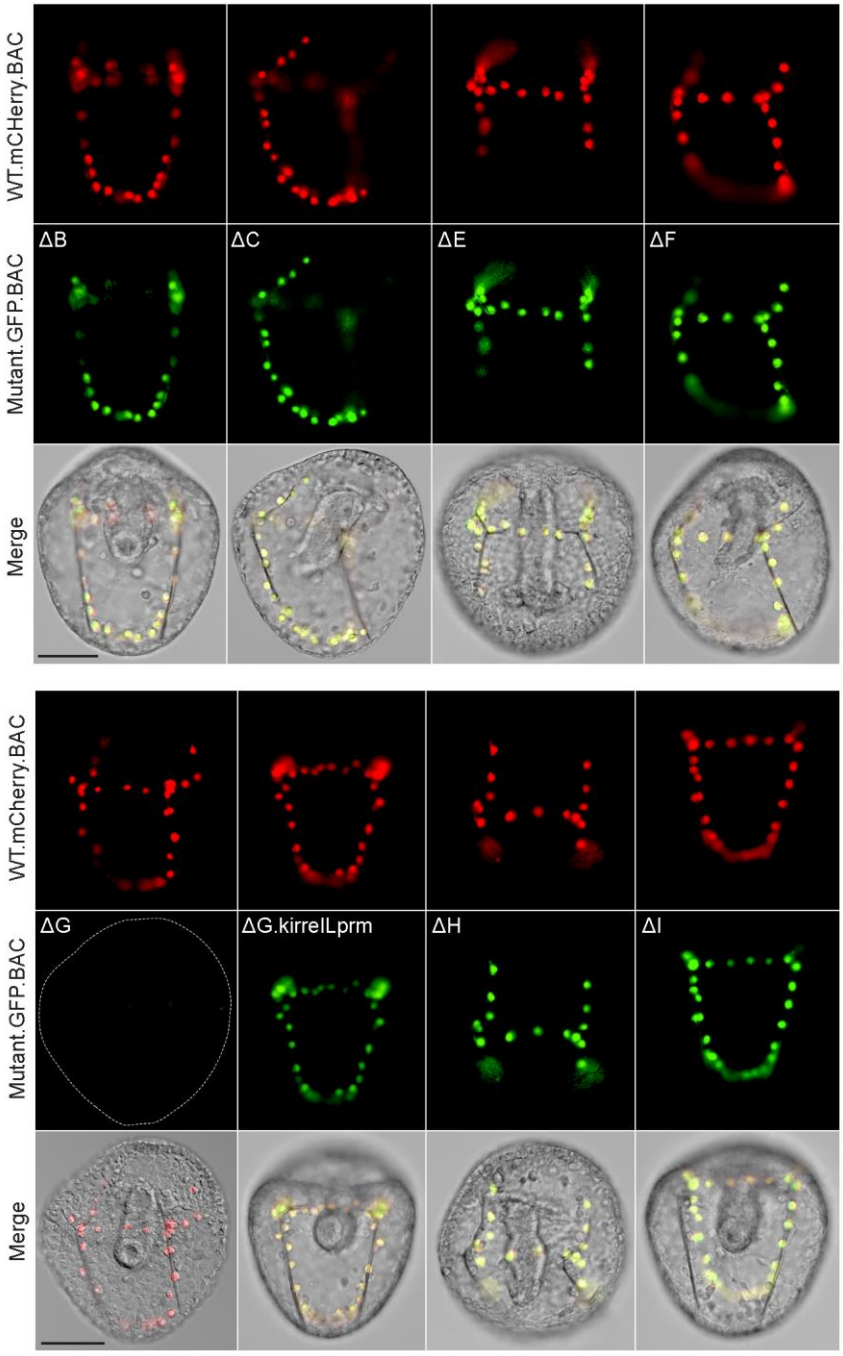
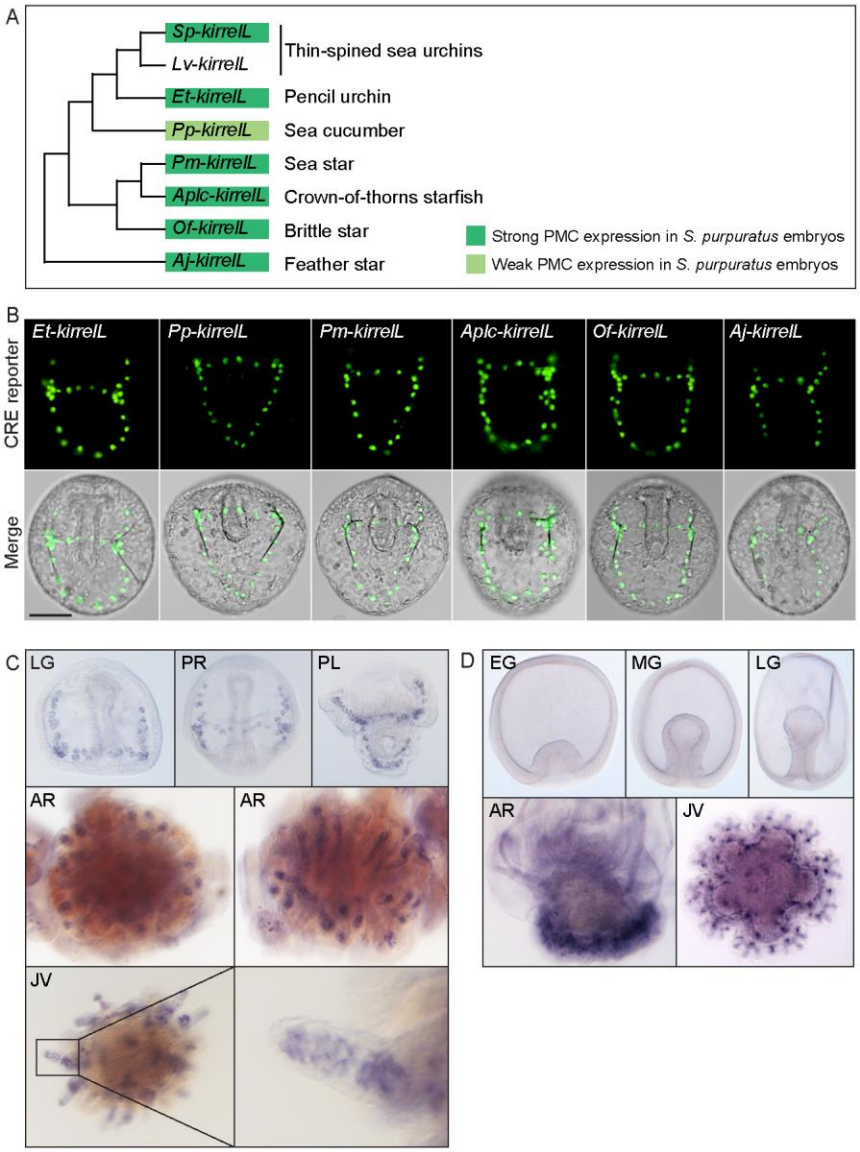


Figure S5: Spatial expression patterns of embryos co-injected with parental mCherry and mutant GFP BACs. Top rows: mCherry fluorescence. Middle rows: GFP fluorescence. Bottom rows: mCherry and GFP fluorescence overlayed onto DIC images. Scale bar: 50 μm.

Cross-species analysis of echinoderm *kirrelL* CREs

As the non-coding region directly upstream of the translational start site of *Sp-kirrelL* was found to contain transcriptional control elements, we asked whether sequences directly upstream of *kirrelL* genes from other echinoderm classes might contain conserved CREs that have activity in *S. purpuratus* PMCs. We cloned non-coding sequences upstream of *kirrelL* genes from *Eucidaris tribuloides* (pencil urchin), *Parastichopus parvimensis* (sea cucumber), *Patiria miniata* (sea star), *Acanthaster planci* (crown-of-thorns starfish), *Ophionereis fasciata* (brittle star), and *Anneissia japonica* (feather star) into the *EpGFPII* plasmid and injected them into fertilized *S. purpuratus* eggs (**Figure 6A**). Remarkably, we found that all six drove GFP expression in sea urchin embryos, with five out of six exhibiting strong GFP expression selectively in PMCs (**Figure 6B**). Taken together, these observations indicate that *kirrelL* CREs across echinoderm species are highly conserved. We found it particularly striking that *kirrelL* CREs from deeply divergent echinoderm species that do not form embryonic or larval skeletons (sea stars and feather stars) drive PMC-selective GFP expression in sea urchin embryos.

Although KirrelL has been shown to be an important morphoeffector gene in the sea urchin embryo, where it plays an essential role in PMC-PMC fusion, its expression in adult sea urchins has not been examined. We observed *Lv-kirrelL* expression in the skeletogenic centers of the adult rudiment and in the spine of five-week-old juvenile sea urchins (**Figure 6B**). The expression pattern of *Lv-kirrelL* was very similar to that of *Lv-msp130r2*, a highly expressed biomineralization gene (**Figure S6B**). In contrast, expression of *Pm-kirrelL* was not detected during early embryonic and larval development in *P. miniata*, which does not from a larval skeleton (**Figure 6C**). *Pm-kirrelL* is, however, expressed in the developing adult rudiment in pre-metamorphic, late-stage sea star larva and in the adult skeletogenic centers in juvenile sea stars (**Figure 6D**). As a control, we showed *Pm-ets1* expression in the mesenchyme cells during early development and an expression pattern in the adult rudiment and skeletogenic centers in juvenile sea stars that closely resembled that of *Pm-kirrelL* (**Figure S6C**).



400

401 **Figure 6**: Cross-species analysis of *kirrelL* CREs from diverse members of the echinoderm
402 phylum. (A) Phylogenetic relationships of *kirrelL* genes based on the consensus view of
403 evolutionary relationships among echinoderms. Branch lengths are not drawn to scale. Box colors
404 correspond to expression of GFP in *S. purpuratus* embryos, driven by non-coding sequences
405 upstream of *kirrelL* genes of *Eucidaris tribuloides* (Et-kirrelL), *Parastichopus parvimensis* (Pp-
406 kirrelL), *Patiria miniata* (Pm-kirrelL), *Acanthaster planci* (Aplc-kirrelL), *Ophionereis fasciata* (Of-
407 kirrelL), and *Anneissia japonica* (Aj-kirrelL). Criteria for strong and weak PMC expression are
408 defined in Figure 2. (B) Spatial expression patterns of GFP reporter constructs containing *kirrelL*
409 CREs from other echinoderm species in *S. purpuratus* embryos at 48 hpf. Top row: GFP
410 fluorescence. Bottom row: GFP fluorescence overlayed onto DIC images. Scale bar: 50 μm. (C)
411 Representative WMISH images showing *Lv-kirrelL* expression during *L. variegatus* development.
412 (D) *Pm-kirrelL* expression during *P. miniata* development. EG, early gastrula; MG, mid-gastrula;
413 LG, late gastrula; PR, prism stage; PL, pluteus larva; AR, adult rudiment; JV, juvenile stage.

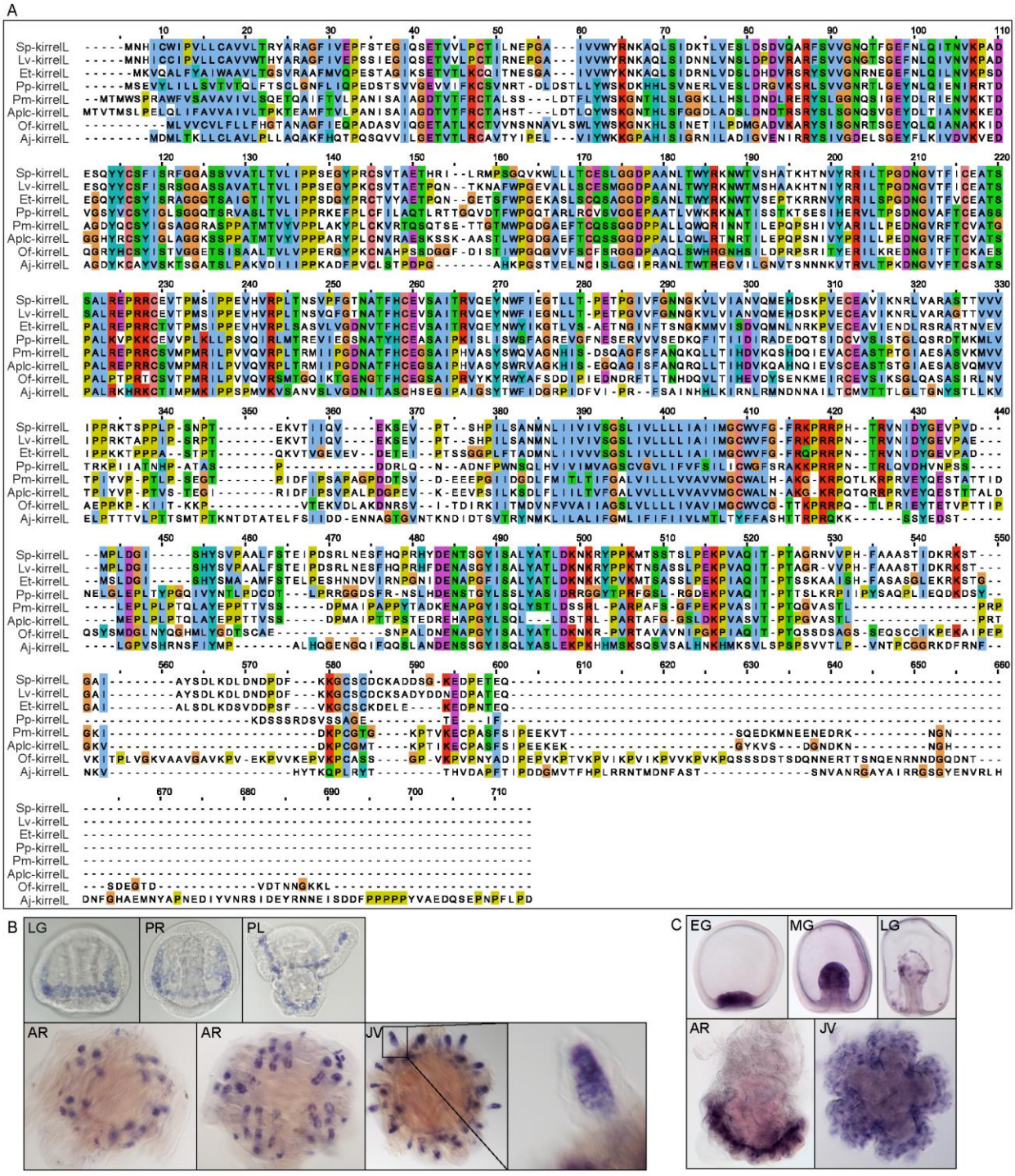


Figure S6: Alignment of echinoderm KirrelL proteins and representative WMISH images of positive control probes. (A) Clustal alignment of echinoderm KirrelL proteins from *Eucidaris tribuloides* (Et-kirrelL), *Parastichopus parvimensis* (Pp-kirrelL), *Patiria miniata* (Pm-kirrelL), *Acanthaster planci* (Aplc-kirrelL), *Ophionereis fasciata* (Of-kirrelL), and *Anneissia japonica* (Aj-kirrelL). Colors correspond to the Clustal default residue coloring scheme. (B) *Lv-msp130r2* expression during *L. variegatus* development. (C) *Pm-ets1* expression during *P. miniata*

development. EG, early gastrula; MG, mid-gastrula; LG, late gastrula; PR, prism stage; PL, pluteus larva; AR, adult rudiment; JV, juvenile stage.

Dissection of a candidate adult skeletogenic CRE

As sea stars do not form a larval skeleton but express *kirrelL* specifically in adult skeletogenic centers, we exploited the activity of the *Pm-kirrelL* CRE in sea urchin embryos as a potential proxy for identifying transcriptional inputs that ordinarily control this gene in adult echinoderms (see Discussion). We performed truncations and mutations of the regulatory regions upstream of the *Pm-kirrelL* gene to identify direct transcriptional inputs (**Figure 7**). Subdivision of the ~4 kb Pm-kirrelL regulatory region showed that activity was restricted to the proximal region (Pm2), and further analysis revealed that a 614 bp region (PmG) that was sufficient to drive strong PMC-specific GFP expression in *S. purpuratus* embryos (**Figure 7B and Figure S7A**). Phylogenetic footpriting of genomic sequences from *P. miniata* and the closely related crown-of-thorns starfish (*A. planci*) showed substantial similarity in this region (**Figure 7A**). We performed mutational analysis of the PmG element and found that this CRE receives positive inputs from both Alx1 and Ets1 (**Figure 7C and Figure S7B**), similar to the *Sp-kirrelL* C and G.ATAC elements. We next asked whether PmG1 and PmG2 elements, which are located near the *Pm-kirrelL* transcriptional start site, could interact with distal *Sp-kirrelL* elements, thereby substituting for the endogenous *Sp-kirrelL* promoter. For this analysis, we generated chimeric *EpGFPII* reporter constructs that contained the sea urchin *Sp-kirrelL* element C (SpC) adjacent to the sea star PmG1 or PmG2 element (**Figure 7D**). We found that PmG1 and PmG2 were both interchangeable with the *Sp-kirrelL* promoter and that interactions between SpC and PmG1 or PmG2 supported strong PMC-specific GFP expression in *S. purpuratus* embryos (**Figure S7C**). In a construct containing a PmG2 element with shuffled sequence, GFP expression was abolished. These observations highlight a striking conservation of sequence and function in *kirrelL* promoters from deeply divergent echinoderm species.

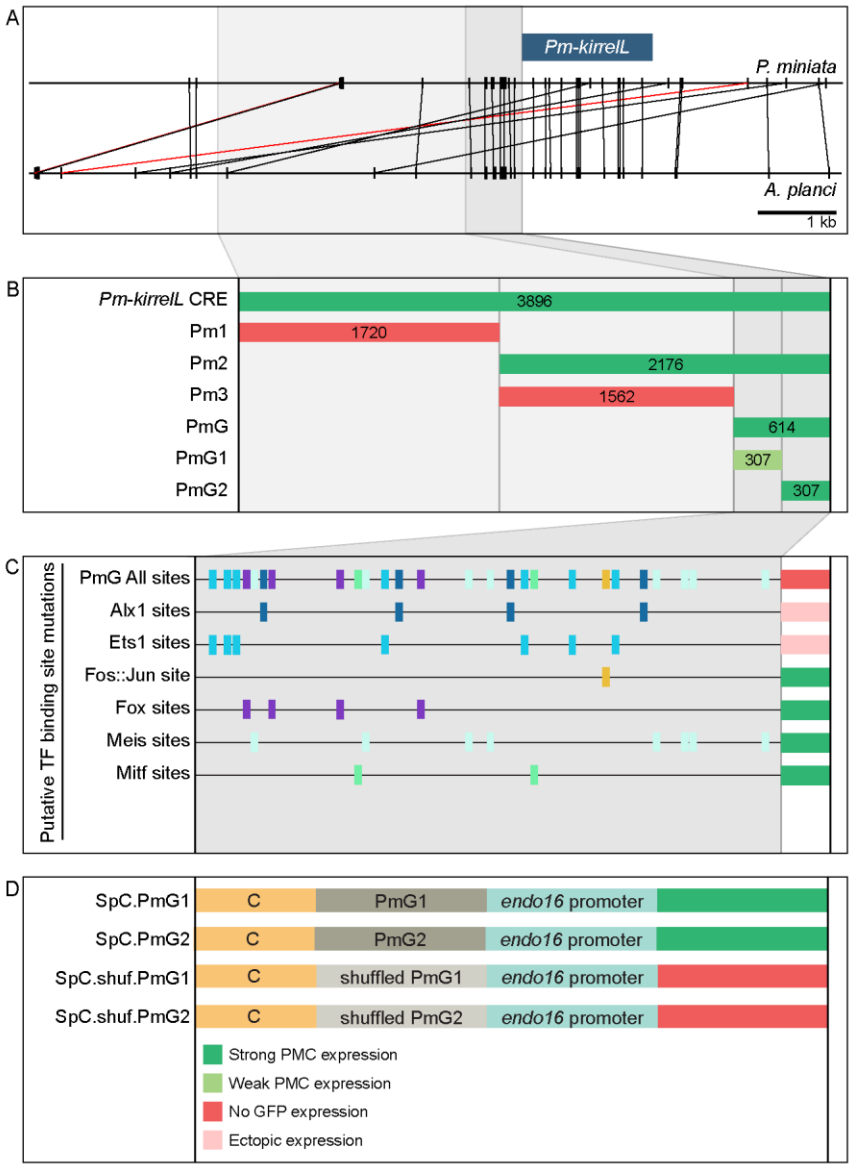


Figure 7: Functional analysis of non-coding genomic sequences upstream of *Pm-kirrelL* to identify CREs. (A) Phylogenetic footprinting of genomic sequences near *P. miniata* and *A. planci* kirrelL using GenePalette. Black lines indicate identical sequences of 15 bp or longer in the same orientation while red lines indicate identical sequences of 15 bp or longer in the opposite orientation. (B) Summary of GFP expression regulated by non-coding sequences upstream of the *Pm-kirrelL* translational start site. (C) Summary of GFP expression driven by PmG element mutants. (D) Summary of GFP expression regulated by chimeric reporter constructs containing *Sp-kirrelL* element C and *Pm-kirrelL* G1 or G2 elements. Criteria for strong and weak PMC expression are defined in Figure 2. Ectopic expression is defined as >10% of total injected embryos expressing GFP, with majority exhibiting expressing in cells other than PMCs.

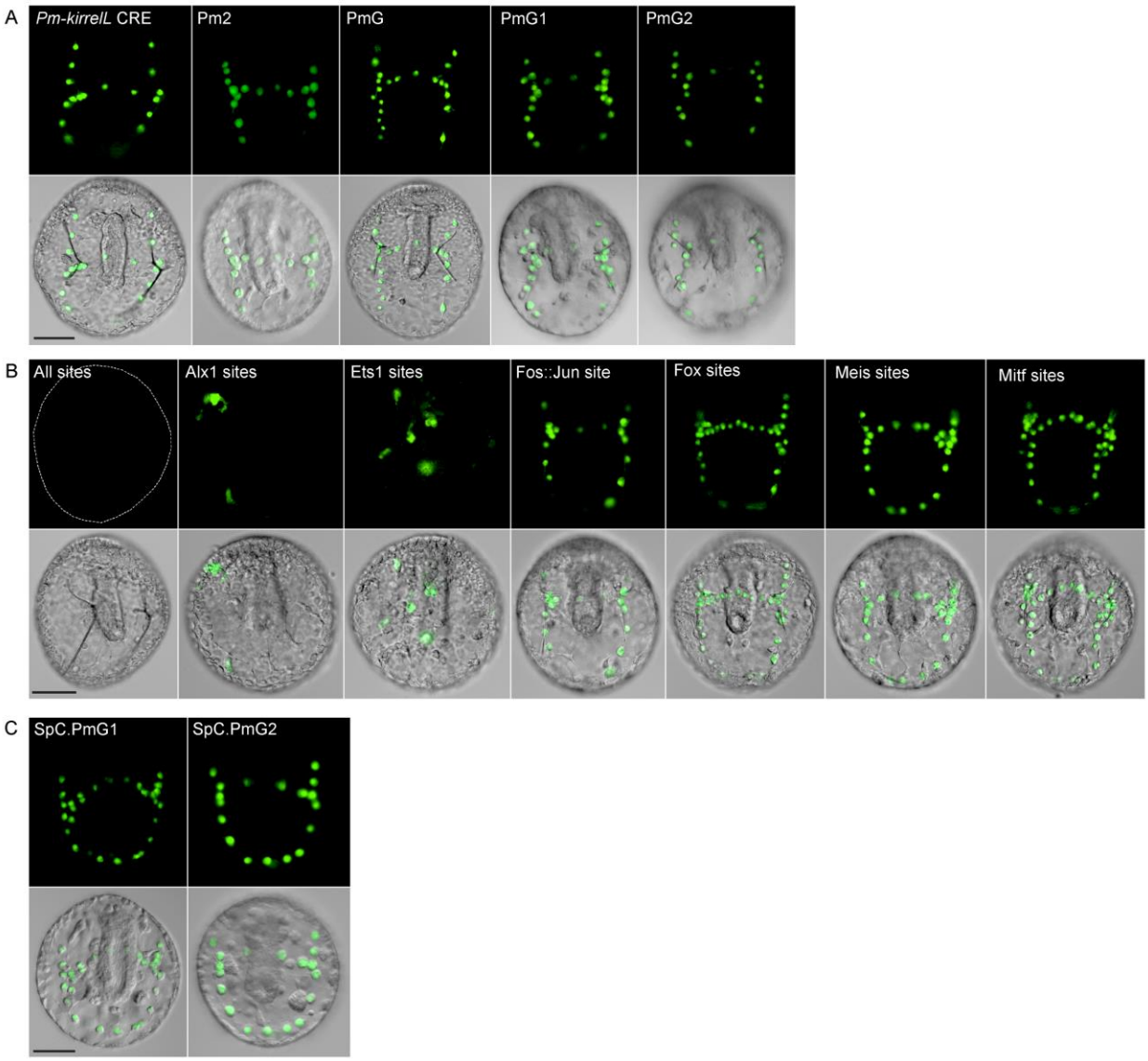


Figure S7: Sea star *Pm-kirrelL* CRE truncation and mutational analysis. (A) Spatial expression patterns of GFP reporter constructs containing *Pm-kirrelL* truncations in *S. purpuratus* embryos at 48 hpf. (B) Spatial expression patterns of GFP reporter constructs containing sea star PmG element mutants in *S. purpuratus* embryos at 48 hpf. (C) Spatial expression patterns of chimeric reporter constructs containing *Sp-kirrelL* element C and *Pm-kirrelL* G1 or G2 elements. Top rows: GFP fluorescence. Bottom rows: GFP fluorescence overlayed onto DIC images. Scale bar: 50 µm.

Discussion

Linking developmental GRNs to morphogenesis

Recent studies with echinoderms have elucidated the architecture of developmental GRNs, including the GRN deployed specifically in embryonic skeletogenic mesenchyme of sea urchins (Shashikant et al., 2018a). Although these studies have focused largely on interactions among regulatory genes that constitute the core of such networks, the importance of GRNs from a developmental perspective is that they underlie the dramatic anatomical changes that characterize embryogenesis (Ettensohn, 2013; Smith et al., 2018). In that respect, GRNs have considerable power in explaining the transformation of genotype into phenotype. Moreover, if GRNs are to be useful in understanding the evolution of morphology, currently a major goal of comparative GRN biology, the developmental mechanisms by which these genetic networks drive morphology must be addressed. This work seeks to partially fill this conceptual gap by elucidating the transcriptional control of *Sp-kirrelL*, an effector gene required for cell-cell fusion, an important morphogenetic behavior of PMCs.

The cis-regulatory apparatus of *Sp-kirrelL*

The combinatorial control of CRE function is important for driving complex gene expression patterns during animal development. In the present study, we identified key regulatory elements and transcription factor inputs that control *Sp-kirrelL* expression. Using plasmid reporter constructs, we identified 7 CREs (elements B, C, E, F, G, H, and I) that were individually sufficient to drive strong PMC-specific GFP expression when placed adjacent to the native *Sp-kirrelL* promoter. Most of these same elements failed to drive reporter expression at detectable levels, however, when cloned directly adjacent to the 140 bp *Sp-endo16* core promoter, a component of EpGFPII, a vector widely used for *cis*-regulatory analysis in sea urchins. As proximal promoter elements have been shown to tether more distal elements in other organisms (Calhoun et al., 2002), we hypothesize that such tethering activity is present in the 301 bp *Sp-kirrelL* promoter element contained in element G. Tethering activity would also account for the fact the regulatory sites in the C element (i.e., those contained in C.ATAC and C.ChIP) must be in close proximity to the *Sp-endo16* promoter to activate transcription, while these same sites can function at a greater distance when working in concert with the endogenous *Sp-kirrelL* promoter. These findings highlight the potential limitations of transgenic reporter assays that rely exclusively on exogenous and/or core promoters.

As multiple CREs were capable of supporting PMC-specific reporter expression in combination with the *Sp-kirrelL* promoter, we performed BAC deletion analysis to determine the relative contributions of these elements to *Sp-kirrelL* expression. We

quantified reporter expression using a newly developed, Nanostring-based assay that allowed us to measure the extent of transgene incorporation and reporter expression. We found that a single deletion of elements A through G entirely abolished GFP expression, even in the presence of the native *Sp-kirrelL* promoter, pointing to this region as the major regulatory apparatus of the gene and demonstrating that any CREs outside this region (including elements H and I) are insufficient to support transcription during embryogenesis. Consistent with plasmid reporter assays, our quantitative BAC analysis confirmed that elements C and G both make major contributions to *Sp-kirrelL* expression. Furthermore, we confirmed that the *Sp-kirrelL* native promoter is required for BAC reporter activity, also consistent with our plasmid reporter assays and with the hypothesis that the CREs are brought into physical contact with the promoter by chromatin looping during transcription. We observed that deletion of element H, which consisted of the *Sp-kirrelL* 3'-UTR, also resulted in decreased expression of the BAC reporter at 30 and 65 hpf. Although an exogenous polyadenylation site was inserted at the 3' end of the reporter coding sequence during BAC recombineering and was therefore present in ftablall constructs, we cannot exclude the possibility that transcription extended beyond this site and that deletion of the 3'-UTR influenced the processing or stability of the *Sp-kirrelL* transcript rather than transcription.

Elements B, E, F, and I each drove PMC-specific reporter expression in plasmid constructs that contained the *Sp-kirrelL* promoter, but deletion of these elements individually from the *Sp-kirrelL* BAC did not quantitatively affect reporter expression at the developmental stages we examined. There are several possible explanations for this. First, these CREs may have no regulatory function *in vivo*. According to this view, the transcriptional activity of these elements in plasmid constructs was an artifact of bringing them in close proximity to the native *Sp-kirrelL* promoter. This view is inconsistent, however, with the fact that most of these elements (B, E, and I) contain other signatures of enhancer activity. All three elements are hyper-accessible in PMCs relative to other cell types at 24 hpf as assayed by ATAC-seq, and elements E and I are also associated with eRNA signal during early development (**Fig. 1**). Moreover, these elements exhibited some degree of promoter specificity in our reporter assays; i.e., they were active in combination with the *Sp-kirrelL* promoter but not the *Sp-endo16* core promoter. These findings suggest that some or all of these elements ordinarily have a regulatory function. They may modulate the precision of *Sp-kirrelL* expression during early development in subtle ways that our assays did not detect (Lagha et al., 2012) or they may be entirely redundant; i.e., deletion of any one of these elements may result in the complete assumption of its function by other elements. This might be the case, for example, if functionally equivalent CREs ordinarily share the *Sp-kirrelL* promoter. Lastly, although these elements are associated with eRNA expression and cell type-specific accessibility early in embryogenesis, their primary function may be to regulate *Sp-kirrelL* expression during stages of development later than those assayed in this study.

Co-regulation of elements C and G by Alx1 and Ets1

The results of both plasmid- and BAC-based reporter assays showed that elements C and G provide crucial inputs into *Sp-kirrelL*. Detailed dissection of these key elements identified consensus Ets1 and Alx1 binding sites that were essential for activity. This finding was consistent with previous evidence that perturbation of *alx1* or *ets1* function using antisense morpholinos results in a dramatic reduction of *Sp-kirrelL* expression (Rafiq et al., 2014). Moreover, ChIP-seq studies have shown that Alx1 binds directly to both elements (Khor et al., 2019). We cannot, however, exclude the possibility that other ETS and homeodomain family members expressed in PMCs (e.g., Erg and Alx4) also bind to these sites. Interestingly, although paired-class homeodomain proteins (including Alx1-related proteins found in vertebrates) are thought to regulate transcription primarily through their binding to palindromic sites that contain inverted TAAT sequences (e.g., ATTANNNTAAT), we identified a half site (ATTA) in element C that was required for activity. This finding supports other recent work which has shown that half sites play a more prominent role in the transcriptional activity of Alx1 than was previously appreciated (Guerrero-Santoro et al., 2021).

Based on gene knockdown studies and the epistatic gene relationships they reveal, Oliveri et al. (2008) proposed that several PMC effector genes are regulated through a feed-forward circuit involving Alx1 and Ets1. They showed that Ets1 positively regulates *alx1* and that both regulatory inputs are necessary to drive expression of several biomineralization-related genes. Our findings support such a model and extend it by demonstrating that the topology of this feed-forward regulation is very simple- both Alx1 and Ets1 provide direct, positive inputs into CREs associated with *Sp-kirrelL*. We identified dual, direct inputs into two different CREs, one associated with the promoter (element G) and a more distal element (element C). Evidence from other recent studies suggest that direct co-regulation by Alx1 and Ets1 is a widespread mechanism for controlling PMC effector gene expression. Genome-wide analysis of Sp-Alx1 ChIP-seq peaks located near effector gene targets showed that both Alx1 and Ets1 consensus binding sites were highly enriched in these regions (Khor et al., 2019) and both Alx1 and Ets1 binding sites are enriched in regions of chromatin that are hyper-accessible in PMCs relative to other cell types (Shashikant et al., 2018b). Our analysis of *Sp-kirrelL* reveals that feed-forward regulation by Alx1 and Ets1 controls not only the expression of biomineralization-related genes but also genes that regulate PMC behavior, thereby integrating these cellular activities.

Davidson (1986) proposed that sea urchins, ascidians, nematodes, and several other animal groups develop by a so-called “Type I” mechanism, a mode of development characterized by the early embryonic expression of terminal differentiation genes. A prediction of this model is that Type I embryos deploy developmental GRNs that are relatively shallow; i.e., there are few regulatory layers between cell specification and cell

584 differentiation. The *cis*-regulatory control of *Sp-kirrelL* by Alx1 and Ets1 supports this
585 prediction; both transcription factors are activated during early embryogenesis and
586 provide direct, positive inputs into *Sp-kirrelL*. Although mutations of other putative
587 transcription factor binding sites in elements C and G did not result in any noticeable
588 effects on reporter expression in our studies, it should be noted that perdurance of GFP
589 mRNA or protein following activation by early regulatory inputs such as Alx1 and Ets1
590 might have masked effects of such mutations on later stages of embryogenesis.

591

592 Evolutionary conservation of echinoderm kirrelL CREs

593 All adult echinoderms have elaborate, calcitic endoskeletons, but larval skeletal
594 elements are found only in echinoids, ophiuroids, and holothuroids (the latter form only a
595 very rudimentary larval skeleton). It is widely believed that the adult skeleton was present
596 in the most recent common ancestor of all echinoderms and that larval skeletons arose
597 subsequently through a developmental re-deployment of the adult program (see reviews
598 by Cary and Hinman, 2017; Koga et al., 2014; Shashikant et al., 2018a). It is debated,
599 however, whether this re-deployment occurred only once, with a subsequent loss of larval
600 skeletons in asteroids, or more than once, with larval skeletons appearing independently
601 in several groups. Our studies establish *kirrelL* as a component of the ancestral
602 echinoderm skeletogenic GRN, which also included *alx1, ets1*, and *vegfr-10-Ig*
603 (Erkenbrack and Thompson, 2019; Shashikant et al., 2018a).

604 There is abundant evidence that mutations in cis-regulatory sequences contribute
605 to phenotypic evolution (Rebeiz and Tsiantis, 2017; Wray, 2007). At the same time, there
606 are examples of evolutionarily conserved GRN topologies and transcription factor binding
607 sites, often between relatively recently diverged taxa (e.g., mice and humans) but
608 sometimes more deeply conserved (Rebeiz et al., 2015). In the present study, we showed
609 that non-coding sequences upstream of the translational start sites of *kirrelL* genes from
610 a diverse collection of echinoderms supported PMC-specific reporter expression in sea
611 urchin embryos. These echinoderms included a crinoid (*A. japonica*) and two sea stars
612 (*A. planci* and *P. miniata*), taxa that diverged from echinoids 450-500 million years ago
613 (Paul and Smith, 1984; Pisani et al., 2012). The deep evolutionary separation of these
614 groups reveals a remarkable conservation of the *kirrelL* regulatory apparatus over this
615 vast time period. Although the amino acid sequences of KirrelL proteins are well
616 conserved within the phylum (**Figure S6A**), the sequences of the upstream regulatory
617 regions we identified are more divergent. Despite limited nucleotide sequence
618 conservation, dissection of the *Pm-kirrelL* regulatory region provided evidence that in sea
619 stars, as in sea urchins, Alx1 and Ets1 provide direct, positive inputs into *kirrelL*.
620 Moreover, we showed that regulatory elements directly upstream of the *Pm-kirrelL*
621 translation start site could substitute for the native *Sp-kirrelL* promoter in supporting the

activity of the *S. purpuratus* C element, an effect that we hypothesize reflects a deep conservation of the binding sites and proteins that mediate CRE-promoter tethering.

The embryonic skeletogenic GRN of sea urchins has been elucidated in considerable detail, but analysis of the ancestral, adult program has this far been limited to comparative gene expression studies, as there are several technical hurdles to molecular perturbations of adult echinoderms. Because sea stars do not express *kirrelL* at embryonic stages and lack a larval skeleton, but express *kirrelL* in adult skeletogenic centers, we conclude that the function of the sea star *kirrelL* cis-regulatory system is to control the transcription of this gene in the adult. Thus, our identification of Alx1 and Ets1 inputs into the *Pm-kirrelL* regulatory region provides evidence that these inputs are required in skeletal cells of the adult sea star, consistent with the finding that both Ets1 and Alx1 are expressed selectively by these cells (Gao and Davidson, 2008). We cannot exclude the possibility that the regulatory interactions we detected in the context of the *S. purpuratus* embryo are vestiges of an ancient, larval skeletogenic program that has since been lost in sea stars, if indeed this was the evolutionary trajectory of larval skeletogenesis within echinoderms. This interpretation, however, would require the evolutionary conservation of the relevant regulatory DNA sequences over a vast period of time despite their complete lack of function, a scenario that seems very unlikely. We propose instead that our findings provide the first glimpse of functional gene interactions in the ancestral, adult echinoderm skeletogenic program and highlight the remarkable conservation of this program in adults and embryos. They strongly support the view that co-option of the adult skeletogenic GRN into the embryo occurred, at least in part, via a heterochronic shift in the expression of Alx1 and Ets1. This would have been sufficient to transfer a large part of the skeletogenic GRN into the embryo, as the transcription of many key effector genes, including *kirrelL*, was already directly linked to Alx1 and Ets1 expression. Direct analysis of CRE structure and function in the adult skeletogenic centers of sea stars and sea urchins will be required to more fully elucidate the architecture of the ancestral network.

Supplemental tables

Supplemental Table S1: Quantification of GFP expression patterns in embryos injected with reporter constructs.

Supplemental Table S2: Summary of NanoString analysis.

Supplemental Table S3: NanoString analysis probe target sequences.

Materials and Methods

Animals

Adult *Strongylocentrotus purpuratus* and *Patiria miniata* were acquired from Patrick Leahy (California Institute of Technology, USA). Adult *Lytechinus variegatus* were acquired from the Duke University Marine Laboratory (Beaufort, NC, USA) and from Pelagic Corp. (Sugarloaf Key, FL, USA). Spawning of gametes was induced by intracoelomic injection of 0.5 M KCl. *S. purpuratus* and *P. miniata* embryos were cultured in artificial seawater (ASW) at 15°C in temperature-controlled incubators while *L. variegatus* embryos were cultured at 19-24°C. Late-stage *L. variegatus* and *P. miniata* larvae were fed with *Rhodomonas lens* algae, accompanied by water changes every other day.

Generation of cis-regulatory reporter constructs

Phylogenetic footprinting between echinoderm kirrelL loci was performed using GenePalette with a sliding window size of 15 bp. GFP reporter constructs were generated by cloning putative CREs into the *EpGFPII* plasmid, which contains the basal promoter of *Sp-endo16* (Cameron et al., 2004). Putative *Sp-kirrelL* CREs were amplified from *S. purpuratus* genomic DNA using primers with restriction site overhangs **(see Supplemental Table S3)**. CREs with mutations of putative transcription factor binding sites and putative CREs from echinoderm species were synthesized as gBlock gene fragments with flanking restriction sites by Integrated DNA Technologies (Coralville, IA, USA). Sequences of putative CREs from echinoderm species (other than sea urchins) were located 2-3 kb upstream of the kirrelL gene translational start sites.

BAC recombineering

Sp-KirrelL BAC-GFP reporter constructs were generated from a parental BAC (R3-28J10-14544) according to established recombineering protocols (Buckley et al., 2018). The recombineering cassettes were synthesized by Integrated DNA Technologies (Coralville, IA, USA). The cassettes contained GFP coding sequence, SV40 terminator sequence, a kanamycin resistance gene between two flippase recognition target (FRT) sites and flanking homologous arms. The recombineering cassettes were transformed into EL250 cells carrying the parental BAC (pBACe3.6 vector harboring *Sp-kirrelL* and flanking genomic sequences) and recombinase genes were de-repressed via heat shock. EL250 cells with recombinant BACs were selected based on kanamycin resistance. To remove the kanamycin resistance gene, expression of *flippase* (*flp*) recombinase enzyme was induced using L-(+)-arabinose and colonies with the kanamycin resistance gene removed were identified by replica plating. BACs without kanamycin resistance gene were subsequently electroporated and propagated in DH10β cells.

DNA microinjection

Microinjection of reporter constructs was performed following established protocols (Arnone et al., 2004). Prior to injection, reporter constructs were linearized and mixed with carrier DNA that was prepared by overnight HindIII digestion of *S. purpuratus* or *L. variegatus* genomic DNA. BAC and plasmid constructs were linearized with AscI and KpnI restriction enzymes, respectively. Each injection solution contained 100 ng/µL linearized DNA, 500 ng/µL carrier DNA, 0.12 M KCl, 20% glycerol, 0.1% Texas Red dextran in DNAse-free, sterile water. *S. purpuratus* embryos were cultured for 48 hpf and *L. variegatus* were cultured for 28 hpf before being mounted for live imaging. Embryos were scored to determine the total number of injected embryos (indicated by the presence of Texas Red dextran), the number of embryos showing PMC-specific GFP expression, the number of embryos showing PMC and ectopic GFP expression, and the number of embryos with only ectopic GFP expression.

NanoString analysis

Direct quantitative measurement of GFP and mCherry RNA transcripts and incorporated DNA was performed using the Nanostring nCounter Elements XT protocol. Briefly, a pair of target-specific oligonucleotide pairs (Probes A and B) complementary to each target gene and transcript were synthesized by Integrated DNA Technologies (Coralville, IA, USA). Probes A and B also included short tails complementary to NanoString Reporter Tags and Universal Capture Tags, respectively. RNA targets included GFP, mCherry, and several *S. purpuratus* housekeeping genes (*foxJ1, hlf, kazL*, and *rasprp3*) that represented a range of transcript abundances and that were expressed at constant levels over the developmental time window of interest. DNA targets included GFP, mCherry, several endogenous, single-copy genes (*hypp_1164, hypp_1901, hypp_2956, hypp_592, kirrelL*), and one multi-copy gene (*pmar1*). DNA probes were complementary to the non-coding DNA strand to avoid hybridization to RNA. Probe sequences are available in **Supplemental Table S3**. For detection, we used the NanoString Elements XT Reporter Tag Set-12 and Universal Capture Tag.

Embryos injected with parental and mutant BACs were harvested at 20, 30, 50 and 65 hpf using the Qiagen AllPrep DNA/DNA micro kit. An additional on-column DNase treatment was included in the RNA recovery process to remove contaminating DNA. Genomic DNA extracted was sonicated using a Bioruptor Pico (Diagenode) for 6 minutes (30 seconds ON, 30 seconds OFF) at 4°C to obtain ~200 bp fragments (confirmed using an Agilent Bioanalyzer). Sonicated DNA was extracted using ethanol precipitation. GFP or mCherry RNA counts were first normalized to housekeeping transcript counts. DNA counts were normalized to single copy gene counts to obtain number of incorporated DNA per nucleus. To obtain RNA count per incorporated DNA for each sample, normalized RNA counts were divided by normalized incorporated DNA counts (**Supplemental Table S2 and S4**).

Whole-mount in situ hybridization

DNA templates for RNA probe synthesis were amplified with reverse primers that contained T3 promoter (**see Table S3**). Invitrogen MEGAscript T3 Transcription Kit was then used to amplify digoxigenin-labeled RNA from the DNA templates. Whole-mount in situ hybridization (WMISH) was performed as previously described (Ettensohn et al., 2007), with minor modifications. Embryos were collected fixed at the desired stage and fixed 4% PFA in artificial seawater (ASW) for 1 hour at room temperature. The embryos were then washed twice in ASW and permeabilized and stored in with 100% methanol. Embryos were then rehydrated and incubated with 1ng/µL RNA probe overnight at 55°C. The following day, the embryos were incubated in blocking buffer (1% BSA and 2% horse serum in PBST) and then in blocking buffer with 1:2000 α-DIG-AP antibody. Excess antibody was washed away and color reaction for alkaline phosphatase was carried out.

741 Amore G, Davidson EH. 2006. cis-Regulatory control of cyclophilin, a member of the ETS-DRI
742 skeletogenic gene battery in the sea urchin embryo. *Developmental Biology* **293**:555–564.
743 doi:10.1016/j.ydbio.2006.02.024
744 Arnone MI, Andrikou C, Annunziata R. 2016. Echinoderm systems for gene regulatory studies in
745 evolution and development. *Current Opinion in Genetics & Development* **39**:129–137.
746 doi:10.1016/j.gde.2016.05.027
747 Arnone MI, Dmochowski IJ, Gache C. 2004. Using Reporter Genes to Study cis-Regulatory
748 ElementsMethods in Cell Biology. Elsevier. pp. 621–652. doi:10.1016/S0091-
749 679X(04)74025-X
750 Baertsch R, Diekhans M, Kent WJ, Haussler D, Brosius J. 2008. Retrocopy contributions to the
751 evolution of the human genome. *BMC Genomics* **9**:466. doi:10.1186/1471-2164-9-466
752 Buckley KM, Dong P, Cameron RA, Rast JP. 2018. Bacterial artificial chromosomes as
753 recombinant reporter constructs to investigate gene expression and regulation in
754 echinoderms. *Briefings in Functional Genomics* **17**:362–371. doi:10.1093/bfgp/elx031
755 Calhoun VC, Stathopoulos A, Levine M. 2002. Promoter-proximal tethering elements regulate
756 enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of*
757 *the National Academy of Sciences* **99**:9243–9247. doi:10.1073/pnas.142291299
758 Cameron RA, Oliveri P, Wyllie J, Davidson EH. 2004. cis-Regulatory activity of randomly chosen
759 genomic fragments from the sea urchin. *Gene Expression Patterns* **4**:205–213.
760 doi:10.1016/j.modgep.2003.08.007
761 Cary GA, Hinman VF. 2017. Echinoderm development and evolution in the post-genomic era.
762 *Developmental Biology* **427**:203–211. doi:10.1016/j.ydbio.2017.02.003
763 Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat*
764 *Rev Genet* **10**:691–703. doi:10.1038/nrg2640
765 Czarkwiani A, Dylus DV, Oliveri P. 2013. Expression of skeletogenic genes during arm
766 regeneration in the brittle star Amphiura filiformis. *Gene Expression Patterns* **13**:464–472.
767 doi:10.1016/j.gep.2013.09.002
768 Davidson EH. 1986. Gene activity in early development. Orlando: Academic Press.
769 Dermody TS, Kirchner E, Guglielmi KM, Stehle T. 2009. Immunoglobulin Superfamily Virus
770 Receptors and the Evolution of Adaptive Immunity. *PLoS Pathog* **5**:e1000481.
771 doi:10.1371/journal.ppat.1000481
772 Dylus DV, Czarkwiani A, Blowes LM, Elphick MR, Oliveri P. 2018. Developmental transcriptomics
773 of the brittle star Amphiura filiformis reveals gene regulatory network rewiring in
774 echinoderm larval skeleton evolution. *Genome Biol* **19**:26. doi:10.1186/s13059-018-1402-
775 8
776 Erkenbrack EM, Thompson JR. 2019. Cell type phylogenetics informs the evolutionary origin of
777 echinoderm larval skeletogenic cell identity. *Commun Biol* **2**:160. doi:10.1038/s42003-
778 019-0417-3
779 Ettensohn CA. 2013. Encoding anatomy: Developmental gene regulatory networks and
780 morphogenesis: Encoding Anatomy. *genesis* **51**:383–409. doi:10.1002/dvg.22380
781 Ettensohn CA, Dey D. 2017. KirrelL, a member of the Ig-domain superfamily of adhesion proteins,
782 is essential for fusion of primary mesenchyme cells in the sea urchin embryo.
783 *Developmental Biology* **421**:258–270. doi:10.1016/j.ydbio.2016.11.006

784 Ettensohn CA, Illies MR, Oliveri P, De Jong DL. 2003. Alx1, a member of the Cart1/Alx3/Alx4
785 subfamily of Paired-class homeodomain proteins, is an essential component of the gene
786 network controlling skeletogenic fate specification in the sea urchin embryo. *Development*
787 **130**:2917–2928. doi:10.1242/dev.00511
788 Ettensohn CA, Kitazawa C, Cheers MS, Leonard JD, Sharma T. 2007. Gene regulatory networks
789 and developmental plasticity in the early sea urchin embryo: alternative deployment of the
790 skeletogenic gene regulatory network. *Development* **134**:3077–3087.
791 doi:10.1242/dev.009092
792 Farré D, Martínez-Vicente P, Engel P, Angulo A. 2017. Immunoglobulin superfamily members
793 encoded by viruses and their multiple roles in immune evasion. *Eur J Immunol* **47**:780–
794 796. doi:10.1002/eji.201746984
795 Gao F, Davidson EH. 2008. Transfer of a large gene regulatory apparatus to a new developmental
796 address in echinoid evolution. *Proceedings of the National Academy of Sciences*
797 **105**:6091–6096. doi:10.1073/pnas.0801201105
798 Gao F, Thompson JR, Petsios E, Erkenbrack E, Moats RA, Bottjer DJ, Davidson EH. 2015.
799 Juvenile skeletogenesis in anciently diverged sea urchin clades. *Developmental Biology*
800 **400**:148–158. doi:10.1016/j.ydbio.2015.01.017
801 Guerrero-Santoro J, Khor JM, Açıkbaş AH, Jaynes JB, Ettensohn CA. 2021. Analysis of the DNA-
802 binding properties of Alx1, an evolutionarily conserved regulator of skeletogenesis in
803 echinoderms. *Journal of Biological Chemistry* **297**:100901. doi:10.1016/j.jbc.2021.100901
804 Khor JM, Guerrero-Santoro J, Douglas W, Ettensohn CA. 2021. Global patterns of enhancer
805 activity during sea urchin embryogenesis assessed by eRNA profiling. *Genome Res*
806 gr.275684.121. doi:10.1101/gr.275684.121
807 Khor JM, Guerrero-Santoro J, Ettensohn CA. 2019. Genome-wide identification of binding sites
808 and gene targets of Alx1, a pivotal regulator of echinoderm skeletogenesis. *Development*
809 **146**:dev180653. doi:10.1242/dev.180653
810 Killian CE, Croker L, Wilt FH. 2010. SpSM30 gene family expression patterns in embryonic and
811 adult biomineralized tissues of the sea urchin, Strongylocentrotus purpuratus. *Gene*
812 *Expression Patterns* **10**:135–139. doi:10.1016/j.gep.2010.01.002
813 Koga H, Morino Y, Wada H. 2014. The echinoderm larval skeleton as a possible model system
814 for experimental evolutionary biology: Evolution of Echinoderm Larval Skeleton. *genesis*
815 **52**:186–192. doi:10.1002/dvg.22758
816 Kurokawa D, Kitajima T, Mitsunaga-Nakatsubo K, Amemiya S, Shimada H, Akasaka K. 1999.
817 HpEts, an ets-related transcription factor implicated in primary mesenchyme cell
818 differentiation in the sea urchin embryo. *Mechanisms of Development* **80**:41–52.
819 doi:10.1016/S0925-4773(98)00192-0
820 Lagha M, Bothma JP, Levine M. 2012. Mechanisms of transcriptional precision in animal
821 development. *Trends in Genetics* **28**:409–416. doi:10.1016/j.tig.2012.03.006
822 Logan CY, Miller JR, Ferkowicz MJ, McClay DR. 1999. Nuclear beta-catenin is required to specify
823 vegetal cell fates in the sea urchin embryo. *Development* **126**:345–357.
824 doi:10.1242/dev.126.2.345
825 Lyons DC, Kaltenbach SL, McClay DR. 2012. Morphogenesis in sea urchin embryos: linking
826 cellular events to gene regulatory network states: Sea urchin gastrulation. *WIREs Dev Biol*
827 **1**:231–252. doi:10.1002/wdev.18

Makabe KW, Kirchhamer CV, Britten RJ, Davidson EH. 1995. Cis-regulatory control of the SM50 gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development* **121**:1957–1970.

Mann K, Poustka AJ, Mann M. 2008. The sea urchin (Strongylocentrotus purpuratus) test and spine proteomes. *Proteome Sci* **6**:22. doi:10.1186/1477-5956-6-22

Mann K, Wilt FH, Poustka AJ. 2010. Proteomic analysis of sea urchin (Strongylocentrotus purpuratus) spicule matrix. *Proteome Sci* **8**:33. doi:10.1186/1477-5956-8-33

Mathelier A, Fornes O, Arenillas DJ, Chen C, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **44**:D110–D115. doi:10.1093/nar/gkv1176

Nishimura Y, Sato T, Morita Y, Yamazaki A, Akasaka K, Yamaguchi M. 2004. Structure, regulation, and function of micro1 in the sea urchin Hemicentrotus pulcherrimus. *Dev Genes Evol* **214**:525–536. doi:10.1007/s00427-004-0442-0

Oliveri P, Carrick DM, Davidson EH. 2002. A Regulatory Gene Network That Directs Micromere Specification in the Sea Urchin Embryo. *Developmental Biology* **246**:209–228. doi:10.1006/dbio.2002.0627

Oliveri P, Tu Q, Davidson EH. 2008. Global regulatory logic for specification of an embryonic cell lineage. *Proceedings of the National Academy of Sciences* **105**:5955–5962. doi:10.1073/pnas.0711220105

Paul CRC, Smith AB. 1984. The Early Radiation and Phylogeny of Echinoderms. *Biological Reviews* **59**:443–481. doi:10.1111/j.1469-185X.1984.tb00411.x

Peng CJ, Wikramanayake AH. 2013. Differential Regulation of Disheveled in a Novel Vegetal Cortical Domain in Sea Urchin Eggs and Embryos: Implications for the Localized Activation of Canonical Wnt Signaling. *PLoS ONE* **8**:e80693. doi:10.1371/journal.pone.0080693

Peter IS, Davidson EH. 2015. Genomic control process: development and evolution. London, UK ; San Diego, CA, USA: Academic Press is an imprint of Elsevier.

Pisani D, Feuda R, Peterson KJ, Smith AB. 2012. Resolving phylogenetic signal from noise when divergence is rapid: A new look at the old problem of echinoderm class relationships. *Molecular Phylogenetics and Evolution* **62**:27–34. doi:10.1016/j.ympev.2011.08.028

Rafiq K, Shashikant T, McManus CJ, Ettensohn CA. 2014. Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. *Development* **141**:2542–2542. doi:10.1242/dev.112763

Rebeiz M, Patel NH, Hinman VF. 2015. Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development. *Annu Rev Genom Hum Genet* **16**:103–131. doi:10.1146/annurev-genom-091212-153423

Rebeiz M, Tsiantis M. 2017. Enhancer evolution and the origins of morphological novelty. *Current Opinion in Genetics & Development* **45**:115–123. doi:10.1016/j.gde.2017.04.006

Richardson W, Kitajima T, Wilt F, Benson S. 1989. Expression of an embryonic spicule matrix gene in calcified tissues of adult sea urchins. *Developmental Biology* **132**:266–269. doi:10.1016/0012-1606(89)90222-4

Shashikant T, Khor JM, Ettensohn CA. 2018a. From genome to anatomy: The architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. *genesis* **56**:e23253. doi:10.1002/dvg.23253

Shashikant T, Khor JM, Ettensohn CA. 2018b. Global analysis of primary mesenchyme cell cis-regulatory modules by chromatin accessibility profiling. *BMC Genomics* **19**:206. doi:10.1186/s12864-018-4542-z

Smith AF, Posakony JW, Rebeiz M. 2017. Automated tools for comparative sequence analysis of genic regions using the GenePalette application. *Developmental Biology* **429**:158–164. doi:10.1016/j.ydbio.2017.06.033

Smith SJ, Rebeiz M, Davidson L. 2018. From pattern to process: studies at the interface of gene regulatory networks, morphogenesis, and evolution. *Current Opinion in Genetics & Development* **51**:103–110. doi:10.1016/j.gde.2018.08.004

Sun Z, Ettensohn CA. 2014. Signal-dependent regulation of the sea urchin skeletogenic gene regulatory network. *Gene Expression Patterns* **16**:93–103. doi:10.1016/j.gep.2014.10.002

Tu Q, Cameron RA, Davidson EH. 2014. Quantitative developmental transcriptomes of the sea urchin Strongylocentrotus purpuratus. *Developmental Biology* **385**:160–167. doi:10.1016/j.ydbio.2013.11.019

Walters J, Binkley E, Haygood R, Romano LA. 2008. Evolutionary analysis of the cis-regulatory region of the spicule matrix gene SM50 in strongylocentrotid sea urchins. *Developmental Biology* **315**:567–578. doi:10.1016/j.ydbio.2008.01.007

Weitzel HE, Illies MR, Byrum CA, Xu R, Wikramanayake AH, Ettensohn CA. 2004. Differential stability of β-catenin along the animal-vegetal axis of the sea urchin embryo mediated by dishevelled. *Development* **131**:2947–2956. doi:10.1242/dev.01152

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**:206–216. doi:10.1038/nrg2063